Ottobre 1990
Numero 266
Anno XXIII
Volume XLV

Il terremoto dell'Irpinia: dieci anni di ricerche

Lo studio degli effetti geologici del terremoto del 1980 e di quelli che lo hanno preceduto fornisce una chiave di lettura unica per la comprensione dei processi sismogenetici che hanno luogo nell'Appennino meridionale

di Enzo Boschi, Daniela Pantosti e Gianluca Valensise

Sono trascorsi 10 anni da uno dei più disastrosi terremoti che hanno colpito l'Italia nel corso dell'ultimo secolo. Con una magnitudo sismica (M_s) di 6,9, comparabile con quelle stimate per il terremoto di Messina del 1908 (M_s 7,1) e per il terremoto di Avezzano del 1915 (M_s 7,0), il terremoto del 23 novembre 1980 ha colpito duramente una vasta regione della Campania e della Basilicata. L'area di massimo danneggiamento, dove gli effetti del terremoto hanno raggiunto un'intensità del X grado della scala MCS (Mercalli-Cancani-Sieberg), è compresa tra le alte valli dei fiumi Ofanto e Sele.

Alcuni paesi completamente rasi al suolo, come Laviano e Conza della Campania, un cospicuo numero di altri paesi gravemente colpiti nel cuore dei propri centri storici e patrimoni artistici, circa 3000 vittime e un enorme numero di feriti rappresentano il tragico bilancio di questo evento. Un mesto scenario che in questa regione dell'Appennino meridionale si è ripetuto molte volte nel corso degli ultimi 2000 anni. Eppure, paradossalmente, è proprio questa tendenza dei grandi terremoti a ripetersi in modo cadenzato, negli stessi luoghi e con le stesse modalità, che ci consente di essere ottimisti sulla possibilità di difenderci dai terremoti di domani.

La condizione necessaria per il raggiungimento di questo obiettivo è però la possibilità di studiare grandi terremoti che permettano ai sismologi di approfondire la conoscenza della struttura interna della Terra, dei processi sismogenetici e della risposta dell'ambiente alle sollecitazioni sismiche. Infatti, a differenza di quanto si verifica per molte delle scienze fisiche e a somiglianza di quanto invece avviene per la gran parte di quelle geofisiche, il sismologo incontra enormi difficoltà anche solo nell'avvicinarsi all'oggetto della propria ricerca. Egli non può, come il suo collega fisico teorico, decidere un esperimento, effettuarlo in un arco di tempo conveniente ed eventualmente ripeterlo per mostrare ad altri colleghi i risultati ottenuti e verificarne la stabilità. Al contrario, al sismologo si richiede grande tempismo e attitudine a trarre vantaggio dagli esperimenti a cui la Terra ci fa assistere senza preavviso, e capacità di ricavare leggi di validità universale dall'osservazione limitata e per lo più incompleta di fenomeni particolari. Così, mentre le altre scienze fisiche hanno progredito quasi esclusivamente in virtù dell'intelligenza e della tenacia dei ricercatori, la sismologia, almeno ai suoi primordi, ha progredito «a salti», e molte delle scoperte fondamentali sono state propiziate o comunque legate al verificarsi di un grande terremoto.

Esempi illustri sono quelli di John Michell che, prendendo le distanze dalle tradizionali vedute dei contemporanei, descrisse in termini modernamente scientifici gli effetti dell'immane terremoto di Lisbona del 1755; di Robert Mallet, che, al ritorno da un viaggio di due mesi nelle zone colpite dal grande terremoto lucano del 16 dicembre 1857, pose i fondamenti della moderna sismologia osservativa coniando egli stesso il nome della nuova disciplina (dal greco *seismós*, scuotimento e *logía*, dottrina); e di Harry Fielding Reid, che utilizzò osservazioni strumentali e di terreno del grande terremoto di San Francisco del 1906 per elaborare la teoria dell'*elastic rebound* (rimbalzo o riassestamento elastico), su cui è fondata la moderna descrizione della sorgente sismica.

Oggi fortunatamente le cose sono cambiate, e anche se la sismologia continua a rappresentare per molti versi la meno oggettiva tra le scienze fisiche, il numero di osservazioni e di dati raccolti su grandi terremoti moderni ha consentito alla comunità scientifica di elaborare modelli della sismogenesi che riproducono con soddisfacente approssimazione la realtà dei processi geologici. Tali modelli, per fare un esempio, hanno permesso a diversi gruppi di ricercatori di prevedere indipendentemente i principali parametri di sorgente, ma non l'istante di accadimento, del terremoto che avrebbe colpito San Francisco e la California centrale il 18 ottobre 1989. L'accuratezza di questa previsione è stata tale da consentire ad amministratori e popolazione di identificare e adeguare le infrastrutture a rischio e quindi giungere sostanzialmente preparati all'appuntamento con il terremoto.

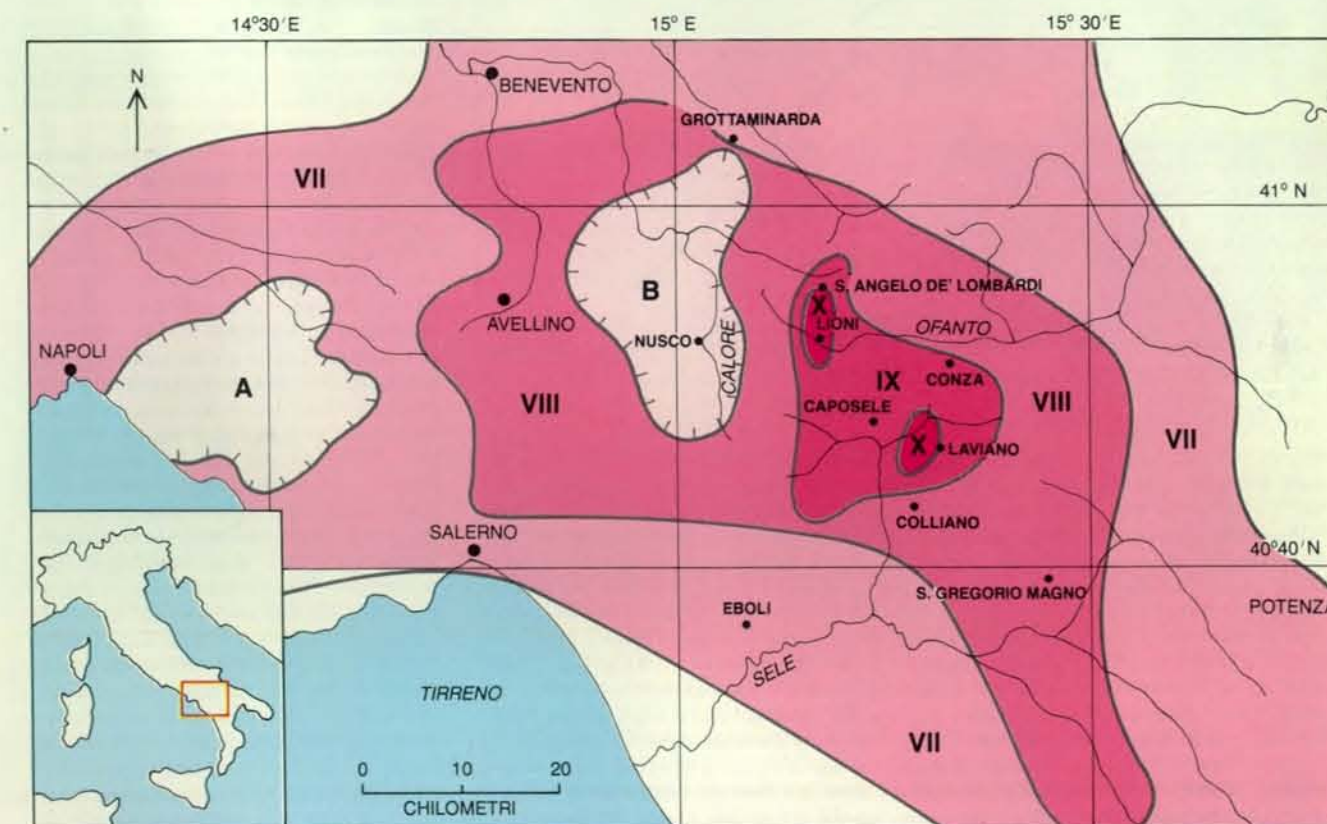
L'esempio di San Francisco indica con chiarezza che dalla sismologia delle soglie del 2000 non ci aspettiamo scoperte rivoluzionarie - come pure potrebbe essere la tanto agognata capacità di prevedere l'istante di accadimento di futuri grandi terremoti - quanto piuttosto la progressiva chiarificazione dei processi che determinano il dove e il come essi avranno luogo e la probabilità che si verifichino entro un determinato intervallo di tempo.

In questo senso il terremoto del 1980, sia pur nella sua drammaticità, ha fornito una opportunità unica per una approfondita comprensione dei processi sismogenetici attivi nella nostra penisola e in particolare nell'Appennino meridionale. Per il fatto di essersi verificato in un momento di piena espansione della sismologia strumentale, il terremoto ha potuto essere studiato in modo moderno e completo. Il moto del suolo indotto dalla scossa principale è stato registrato in un ampio spettro di frequenze e distanze focali, da strumenti a bassa sensibilità espressamente progettati per registrare in zona epicentrale i forti movimenti del suolo e da strumenti sensibilissimi sparsi in tutto il globo. Una rete sismometrica portatile, prontamente installata nella regione colpita a cura di un consorzio di istituzioni italiane, inglesi e francesi, ha consentito di seguire l'evoluzione del fenomeno sismico attraverso la registrazione e la localizzazione di migliaia di repliche.

La quantità di dati sismometrici resisi disponibili fin dai primi mesi del 1981 ha immediatamente qualificato il terremoto del 1980 come l'evento sismico meglio documentato tra quelli avvenuti in Europa. Uno degli aspetti di maggior interesse è stata la possibilità di osservarne e documentarne in maniera oggettiva gli esiti permanenti, in forma di modificazioni della topografia e alterazioni del regime delle acque di superficie e sotterranee verificatesi nella zona epicentrale. Il significato geodinamico di queste osservazioni e la possibilità di utilizzarle per previsioni sull'attività sismica futura della regione formano appunto l'oggetto del nostro studio.

La sorgente sismica e la sua descrizione

È noto che un terremoto ha luogo quando un volume di roccia, sottoposto a tensione per effetto di una coppia di forze di segno opposto guidate a loro volta da forze tettoniche a scala continentale, si rompe dando luogo a scorrimento lungo un piano comunemente indicato come faglia.

Il moto lungo la faglia scarica istantaneamente le tensioni accumulatesi fino al raggiungimento del carico di rottura caratteristico del particolare tipo di roccia, e l'istante successivo apre un nuovo ciclo di accumulo e rilascio. Da questo momento il piano di faglia originatosi nel primo terremoto diventerà il punto preferenziale di rilascio degli sforzi accumulati e la sua superficie verrà a identificarsi in modo sempre più netto a seguito di ognuno degli eventi successivi. Questi processi hanno luogo nella porzione più

Il terremoto del 23 novembre 1980 ha colpito una vasta regione dell'Appennino campano-lucano. Qui sono riportate le isosisme, o linee di ugual intensità degli effetti del terremoto, tracciate a cura dei ricercatori del Progetto finalizzato «Geodinamica» del CNR in base a una dettagliata ispezione dei danni subiti dai centri abitati. I valori di intensità sono espressi nella scala MCS (Mercalli-Cancani-Sieberg). L'area di massimo danneggiamento (*in rosso più scuro*), in cui si è raggiunta un'intensità del IX-X grado della scala MCS, è localizzata in corrispondenza delle alte valli dei fiumi Sele e Ofanto, dove molti centri abitati sono stati totalmente distrutti e numerosissime sono state le vittime. Le tonalità di colore più chiare nelle aree circostanti indicano zone di intensità progressivamente decrescente, ma ancora caratterizzate da estesi danni strutturali. Si noti come l'area del VII grado MCS presenti una lobatura anomala verso la penisola sorrentina e verso Napoli. In corrispondenza delle «isole» *A* e *B*, la prima coincidente con l'area vesuviana, la seconda posta nell'alta valle del fiume Calore, l'intensità raggiunta è stata anche di due gradi inferiore a quella dell'area in cui sono incluse. Fenomeni di questo tipo sono in genere legati alla natura della crosta al di sotto della zona interessata, mentre variazioni di intensità a scala locale, per esempio tra un quartiere e l'altro dello stesso centro abitato, vanno ricercate nella natura delle rocce affioranti.

superficiale e fredda della crosta terrestre, il cosiddetto *strato sismogenetico*, che generalmente si estende fino a 10-15 chilometri di profondità. In condizioni particolari, come lungo le cosiddette zone di subduzione, si possono avere grandi terremoti a profondità di centinaia di chilometri ma, proprio per questa caratteristica, l'impatto che tali eventi hanno sull'ambiente è in generale trascurabile.

L'ipocentro, che i sismologi localizzano analizzando sismogrammi registrati presso diversi osservatori, rappresenta in effetti il punto di nucleazione del processo di frattura, ovvero il punto da cui le rocce poste ai due lati della faglia iniziano a scorrere le une rispetto alle altre coinvolgendo una superficie progressivamente più estesa. Tuttavia un punto non è sufficiente a descrivere un piano,

Innumerevoli sono i terremoti che hanno arrecato distruzioni nell'Appennino meridionale. La storia dell'arte, dell'urbanistica e dei grandi cambiamenti demografici dell'area compresa tra l'Abruzzo e la Sicilia è scandita dal verificarsi di eventi distruttivi e dalle modificazioni indotte da questi nella dinamica sociale. Chiese e conventi ricostruiti sulle macerie dei precedenti, villaggi abbandonati e ricostruiti altrove per sconvolgimenti della topografia e del regime delle acque, sostituzione dei toponimi a significare il taglio con il passato testimoniano il ruolo dei terremoti come creatori di storia se non, paradossalmente, come agenti di rinnovamento e progresso sociale. La rilettura dei documenti storici, principalmente provenienti da archivi ecclesiastici, ha consentito di localizzare i principali terremoti storici, di valutarne le dimensioni e di studiarne gli effetti, sintetizzando le informazioni nel cosiddetto «catalogo sismico storico». In generale, l'attendibilità dei dati in esso riportati cresce con le dimensioni dell'evento e decresce con l'aumentare del tempo trascorso dal suo verificarsi. Qui viene mostrata la localizzazione dei principali eventi riportati per il periodo 1000-1980 dal *Catalogo dei terremoti italiani* del Progetto finalizzato «Geodinamica». Le dimensioni del simbolo sono proporzionali agli eventi. Questi si concentrano lungo una dorsale sismogenetica fondamentale che coincide con l'asse della catena appenninica seguendo la direzione NO-SE. La forte concentrazione spaziale di alcuni gruppi di terremoti fa ritenere che si tratti di eventi simili generati dalla medesima struttura. La non perfetta coincidenza degli epicentri può essere spiegata in alcuni casi ipotizzando che eventi di dimensioni intermedie possano aver interessato solo porzioni distinte della struttura; in altri casi occorre tenere conto della disomogeneità delle descrizioni di eventi avvenuti a secoli di distanza. La documentazione storica, congiuntamente a dati di diversa natura, prevalentemente geologici e morfologici, consente di localizzare le strutture sismogenetiche principali lungo le quali, in questa regione dell'Appennino, si generano i grandi eventi sismici. Qui esse sono rappresentate da linee spesse che indicano l'intersezione del piano di faglia con la superficie topografica. Lungo tale piano ha luogo la rottura che genera il terremoto e che si propaga dall'ipocentro fino, in condizioni favorevoli, alla superficie.

e dunque l'informazione sui parametri ipocentrali è solo il primo passo nell'indagine dell'evoluzione di tale processo.

Da un punto di vista sismografico, il passo successivo è in genere rappresentato dallo studio dell'orientazione nello spazio del piano di faglia e del senso dello scorrimento relativo lungo di esso o, in altre parole, della meccanica focale del terremoto. Come Reid osservò lucidamente dopo il terremoto del 1906, ogni punto della regione circostante l'ipocentro del terremoto registra lo «scatto» rappresentato dallo scorrimento delle rocce che si fronteggiano ai due lati della faglia come un impulso che tende a comprimere o a rarefare i materiali costituenti la porzione più superficiale della crosta. La distribuzione dei volumi di roccia in compressione o rarefazione è funzione della posizione rispetto alla faglia secondo le leggi fondamentali della meccanica. Da ciò consegue che, disponendo di un certo numero di sismogrammi registrati secondo azimut diversi rispetto all'ipocentro del terremoto, è possibile identificare la geometria del piano di faglia e il senso del movimento verificatosi lungo di esso, e in ultima analisi comprendere la dinamica degli sforzi tettonici in atto.

In regioni oggi sottoposte a estensione, come la zona del *rift* del Mar Rosso o come il nostro Appennino centromeridionale, i terremoti più grandi hanno luogo lungo faglie distensive (o «normali»). In regioni sottoposte a compressione e raccorciamento, come nell'area dell'Himalaya per effetto della spinta in senso meridiano impressa dal subcontinente indiano, i terremoti più grandi dimostrano la tendenza di un volume a sormontare quello a esso adiacente lungo faglie che, per contrapposizione alle precedenti, vengono definite «inverse». Infine, laddove grandi zolle crostali scorrono le une vicine alle altre senza significative variazioni della quota relativa, come lungo la grande Faglia di San Andreas in California o lungo la Faglia anatolica in Turchia, si avranno faglie «trascorrenti». Nella realtà fisica le faglie, rispondendo alla somma di sforzi di diversa orientazione e intensità, si muoveranno secondo una combinazione degli stili descritti, per esempio con una modesta componente inversa associata a una cospicua componente orizzontale come durante il recente terremoto di San Francisco.

La conoscenza della geometria del piano di faglia, della sua localizzazione e della direzione del movimento relativo lungo di esso rappresenta il fondamento della descrizione del terremoto sotto il profilo geodinamico. Tuttavia due aspetti fondamentali - la quantità di energia emessa dalla sorgente e le modalità della sua liberazione - devono ancora essere esplorati prima che le conoscenze accumulate possano essere utilizzate nella pratica per creare una modellistica di eventi futuri. Entrambi questi

Un terremoto ha luogo quando un volume di roccia, sottoposto a tensione, si rompe dando luogo a scorrimento relativo lungo un piano di debolezza, detto piano di faglia. Questi processi avvengono nella porzione più superficiale e fredda della crosta terrestre, il cosiddetto «strato sismogenetico», il cui spessore è generalmente di 10-15 chilometri. La rottura ha inizio nell'ipocentro (*asterisco*) e si propaga lungo il piano di faglia, consentendo in tal modo lo scorrimento relativo dei due lembi della faglia. La geometria della faglia e il senso del movimento lungo di essa saranno determinati dalla combinazione di sforzi tettonici a carattere locale e regionale. La figura sintetizza anche le tre classi fondamentali in cui le faglie vengono suddivise in relazione al tipo di movimento che ha luogo lungo di esse.

aspetti sono intimamente legati alle dimensioni del piano di faglia associato a un dato terremoto e all'entità dello scorrimento che vi ha avuto luogo. Queste grandezze vengono combinate linearmente in una grandezza scalare la cui denominazione, «momento sismico», richiama alla mente l'azione delle due forze di segno opposto che determinano il terremoto. Essa va a integrare i tradizionali indicatori delle dimensioni di un terremoto quali l'intensità, ovvero la valutazione degli effetti dello scuotimento sulle cose e sulle persone, e la magnitudo, una grandezza proporzionale all'ampiezza sul sismogramma di particolari onde irraggiate dalla sorgente sismica.

La valutazione oggettiva della gravità di un terremoto è stata per secoli il problema fondamentale dell'osservazione sismologica. La storia riporta che, a seguito del grande terremoto della Capitanata del 1627, venne tracciata la prima carta isosismica, una mappa in cui le località che avevano sofferto un ugual numero di crolli o di vittime venivano unite fra loro a formare aree circolari di ugual scuotimento. Il tempo e la disponibilità di osservazioni strumentali hanno reso il problema progressivamente più complesso e sfaccettato, ma per i nostri scopi basterà accettare il fatto che né la magnitudo, troppo strettamente legata alle condizioni sperimentali in cui ha luogo la misurazione, né tanto meno l'intensità, influenzata da fattori geografici e sociali del tutto indipendenti dal fenomeno sismico, vengono oggi ritenute indicatori fedeli delle reali dimensioni di un terremoto. Quindi, anche se queste grandezze non possono essere abbandonate in quanto riferimento imprescindibile per studiare la sismicità del passato, esse sono in corso di progressiva sostituzione in favore del momento sismico, il cui significato fisico e la cui oggettività meglio si prestano a una descrizione moderna e completa della sorgente di un grande terremoto.

Grandi terremoti e deformazioni crostali

Questa breve digressione sui modi per stimare le dimensioni di un terremoto ci fornisce lo spunto per focalizzare l'attenzione su un aspetto del fenomeno sismico fino a oggi alquanto poco esplorato: il rapporto che intercorre tra il terremoto visto come fenomeno transitorio e istantaneo, le dimensioni fisiche del volume di roccia che, rompendosi, determina l'evento e l'evoluzione della geologia e topografia nella regione che dall'evento è interessata. Ma per rendere chiara la portata di questo tema è necessario fare un passo indietro.

Ancora ai primordi dell'umanità, il terremoto veniva considerato esclusivamente per i suoi effetti materiali, e l'idea di castigo divino spesso implicita nelle stragi e distruzioni prodotte da eventi sismici del passato ha per lungo tempo prevenuto anche gli eruditi dal dare al fenomeno una spiegazione fisica complessiva. Superati gli ostacoli di natura, per così dire, trascendentale, nel secolo scorso la scienza ha cominciato a descrivere la fisica di vari aspetti del fenomeno sismico, dalla velocità di propagazione delle onde all'attenuazione del moto del suolo con la distanza dal fuoco, ma sempre guardando al terremoto come evento unico e isolato. Finalmente, dalla combinazione della grandiosità dei fenomeni naturali nei neocolonizzati territori dell'Ovest degli Stati Uniti con la genialità di Grove Karl Gilbert, acuto naturalista e geografo, sarebbero nate le premesse per dare al terremoto il suo vero posto nell'ambito della storia del pianeta. Nel 1884, visitando le zone disastrate da un terremoto di magnitudo 7,6 che 12 anni prima aveva colpito la Owens Valley (al confine tra California e Nevada), egli annotò che «... le montagne crescono a poco a poco... formando come una piccola scogliera che i geologi conoscono come scarpata di faglia... . Momenti di coesione e di slittamento caratterizzano questa crescita... . Ci sono zone dove le scarpate chiaramente mancano, la deformazione aumenta lentamente, e un giorno essa vincerà la frizione, innalzerà le montagne di alcuni piedi e ripeterà su una scala ancora più terrificante la catastrofe di Owens Valley... ».

Anche se il linguaggio di Gilbert può apparire a tratti enigmatico, le sue osservazioni testimoniano la lucidissima intuizione di un processo, semplice e grandioso nello stesso tempo, che può essere riassunto in alcuni brevi punti. Innanzitutto, la sorgente di un grande terremoto non è rappresentabile come un punto, ma ha dimensioni fisiche valutabili nell'ordine dei chilometri o decine di chilometri. Per eventi di magnitudo superiore a 6,5 queste dimensioni sono tali da coinvolgere lo strato sismogenetico nella sua interezza; in queste condizioni il piano di faglia giunge a intersecare la superficie topografica determinando la formazione di uno scalino nel terreno, indicato come «scarpata di faglia», il cui andamento rispecchia la geometria del fenomeno che ha luogo in profondità. Se la faglia ha carattere predominante normale o inverso, la scarpata rappresenta anche il confine tra una regione della crosta che è stata sollevata dal terremoto e una che, almeno relativamente alla prima, è sprofondata. Come è confermato da numerosi esempi e giustificato dal punto di vista teorico, questa alterazione della to-

pografia è in larga misura permanente e, anche se la scarpata viene in genere rapidamente demolita dall'azione degli agenti erosivi, la differenza di quota tra i due lati della faglia resta e può essere documentata attraverso semplici misurazioni di livellazione geodetica.

La ripetizione di terremoti come quello che ha prodotto la scarpata, ognuno dei quali libererà le tensioni progressivamente accumulate a partire dall'evento precedente, determinerà la progressiva identificazione di una topografia con valli e dorsali, a cui si sovraimporrà un reticolo fluviale progressivamente più attivo con l'aumentare del contrasto altimetrico. Ripetendo per un migliaio di volte un evento in grado di produrre una scarpata alta un metro si costruirà una topografia che, pur tenendo conto dell'erosione delle creste montuose e del parallelo colmamento dei solchi vallivi adiacenti, sarà caratterizzata da un contrasto di 600-700 metri, un valore paragonabile a quello che si osserva nei settori a più elevata sismicità del bacino mediterraneo.

Le osservazioni di Gilbert hanno influenzato in modo decisivo gli studi sismologici di terreno degli anni successivi. Il semplice processo da lui intuito e da noi sinteticamente descritto è stato esplorato in vari aspetti di dettaglio, e i suoi fondamenti ripetutamente verificati e corroborati da osservazioni provenienti da varie parti del mondo.

La comprensione dei meccanismi che controllano l'evoluzione della geologia e della topografia in una regione a elevata attività tettonica fornisce alle scienze della Terra uno di quegli strumenti di cui i geologi hanno bisogno per comprendere eventi verificatisi in un passato molto più remoto e registrati da rocce antichissime. Ma soprattutto, e su questo aspetto focalizzeremo la nostra attenzione, le intuizioni di Gilbert hanno aperto la strada allo studio dell'attività sismica presente, passata e forse futura, mediante gli strumenti della classica geologia di superficie.

L'enorme potenziale di questi strumenti è divenuto particolarmente chiaro nell'ambito del grande tema rappresentato dalla valutazione del rischio sismico associato a una regione, a un insediamento o a grandi opere civili quali ponti, dighe, centrali elettriche. Per fare un esempio, in particolari condizioni geologiche e climatiche, come quelle che si verificano normalmente negli Stati Uniti occidentali, il rilevamento di dettaglio di scarpate di faglia e di fenomeni di alterazione della topografia a esse associati ha consentito di valutare con successo il potenziale sismogenetico di regioni abitate da un tempo troppo breve per consentire l'identificazione per via storica delle sorgenti sismiche attive. I risultati di queste ricerche hanno permesso a loro volta di valutare l'estensione delle aree ad alto rischio prescindendo da alcune delle principali limitazioni e ambiguità interpretative che caratterizzano i dati sulla sismicità storica.

Probabilmente, il più brillante tra i risultati conseguiti nello studio dei grandi terremoti attraverso l'approccio geologico è l'aver constatato che essi hanno una fondamentale tendenza a replicare se stessi lungo le medesime faglie. Questa constatazione, verificata ormai in tutte le ambientazioni tettoniche conosciute, implica che gli stessi segmenti di più estese strutture sismogenetiche tenderanno a rompersi ripetutamente dando luogo a uno scorrimento di entità pressoché costante, che i confini tra un segmento e il successivo rimarranno stazionari nel tempo e che, come logica conseguenza del fatto che alla scala dei millenni la velocità dei grandi movimenti tettonici può essere assunta costante, i terremoti così generati tenderanno a ricorrere con regolarità. La teoria del «terremoto caratteristico», sviluppata da David Schwartz dello US Geological Survey e da Kevin Coppersmith, oggi consulente per la Geomatrix di San Francisco, sulla base di questa constatazione, ha quindi come obiettivo fondamentale il riconoscimento di un terremoto «campione» per ogni segmento di una specifica struttura sismogenetica. I caratteri di questo campione costituiscono la miglior previsione possibile sui caratteri del prossimo grande evento generato lungo tale segmento. Inoltre l'individuazione di terremoti caratteristici avvenuti lungo segmenti adiacenti a quello a cui si è interessati potrà portare a valutazioni soddisfacenti anche in condizioni di scarsità o totale assenza di informazione, sia storica sia geologica. E ancora, e questo rappresenta forse l'aspetto di interesse pratico più immediatamente comprensibile, l'assunzione di regolarità della comparsa dei terremoti caratteristici implica che, una volta determinato l'intervallo di tempo che mediamente intercorre tra due eventi simili successivi (detto «tempo di ricorrenza»), saremo in grado di esprimere valutazioni non solo sul *dove* e sul *come*, ma anche sul *quando* si verificherà il prossimo terremoto catastrofico.

Tuttavia, il coinvolgimento del fattore tempo sembra andare oltre le possibilità dell'osservazione geologica di terreno. Come già Gilbert aveva notato chiaramente, per costruire una topografia che risalti con evidenza all'occhio del geologo sono necessari molti terremoti in un arco di tempo la cui durata è molto difficile da valutare. Anche conoscendo con precisione l'età delle rocce costituenti una dorsale montuosa che sappiamo essere stata generata dall'attività di una faglia ben individuata, nella maggior parte dei casi ci accorgeremo che esse presentano tracce di cicli di deformazione ormai estinti e che pertanto sono molto più antiche dei fenomeni deformativi a cui assistiamo oggi. Né potrebbe essere d'aiuto lo studio della velocità con cui la dorsale viene smantellata dai processi erosivi, per l'enorme variabilità e dipendenza da fattori climatici transitori a cui questi sono soggetti. Così, dopo molti decenni durante i quali numerose faglie attive erano state accuratamente cartografate e descritte con cognizioni largamente approssimative dell'andamento nel tempo degli eventi sismici, gli ultimi 20 anni hanno visto l'affermarsi di una nuova disciplina che, imponendo stringenti vincoli cronologici ai processi naturali fino a ora descritti, rende finalmente quantitative le osservazioni di Gilbert e dei suoi successori.

La paleosismologia

Questa disciplina, a cui è stato imposto il nome di «paleosismologia», o scienza dei terremoti antichi, utilizza la «lettura geologica» degli eventi sismici come chiave di interpretazione per la comprensione di quelli futuri. Essa ha ricevuto un forte impulso negli Stati Uniti e in Giappone a partire dall'inizio degli anni settanta, in relazione alla crescente esigenza di valutare con confidenza il rischio sismico legato a una veloce urbanizzazione e alla progettazione di grandi opere di ingegneria civile. In particolari regioni la combinazione della assoluta mancanza di osservazioni storiche con l'aridità del clima e con lo scarso grado di urbanizzazione, elementi questi

Il terremoto del 23 novembre 1980 ha avuto magnitudo 6,9. Il suo ipocentro è stato localizzato a 40 gradi e 46 primi di latitudine nord, 15 gradi e 20 primi di longitudine est e a una profondità di 12-14 chilometri. L'epicentro, ovvero la proiezione in superficie dell'ipocentro (indicato dall'asterisco), ricade presso l'abitato di Laviano, uno dei centri più colpiti. Il terremoto è stato generato da un regime tettonico che produce estensione in direzione ortogonale alla faglia. A partire dall'ipocentro la rottura si è propagata lungo il piano di faglia, che è risultato composto di quattro frammenti, per una lunghezza totale di circa 50 chilometri. La proiezione in superficie di tali frammenti è rappresentata dalle aree ombreggiate. Per ognuno di questi frammenti, a eccezione di quello localizzato a nord-est dell'epicentro, lo scorrimento in profondità si è manifestato in superficie come una scarpata di faglia (*in tratto spesso con i dentini che indicano il lato ribassato*) la cui altezza arriva anche a un metro. La sezione in basso a sinistra mostra la geometria della struttura in profondità. Nei mesi successivi al terremoto migliaia di altre scosse di assestamento, solo in parte avvertibili dalla popolazione, hanno interessato l'area epicentrale (*in colore*). Oltre alla scarpata di faglia, lo scorrimento ha prodotto variazioni altimetriche permanenti dell'ordine delle decine di centimetri in una regione di quasi 1000 chilometri quadrati. Dalla distribuzione di queste deformazioni è possibile localizzare i frammenti di faglia, individuarne la geometria e valutare l'entità dello scorrimento che il terremoto ha determinato lungo ognuno di essi. Le isolinee quantificano la variazione di quota (in metri) che ci si può attendere dal modello di fagliazione così elaborato. Il complesso delle osservazioni di superficie consente di esprimere le dimensioni dell'evento in termini di momento sismico. Moltiplicando l'area complessiva della faglia per l'entità media dello scorrimento e per il modulo medio di rigidità delle rocce nella regione si ottiene così un valore di $1{,}8 \times 10^{26}$ dyne per centimetro, che è in buon accordo con le stime ottenute per via sismometrica.

Le faglie che vengono attivate durante i terremoti più grandi, per esempio tutti quelli di magnitudo superiore a 6,5, hanno dimensioni tali da coinvolgere interamente lo strato sismogenetico, giungendo a interessare la superficie topografica. L'emergenza in superficie del piano di faglia determina la formazione di una scarpata di faglia, che riflette in pieno la geometria e la dinamica del fenomeno che ha luogo in profondità. Il terremoto del 23 novembre 1980 ha prodotto una scarpata con orientazione NO-SE ed estesa per quasi 40 chilometri tra Lioni e il Pantano di San Gregorio Magno. La scarpata mostra una sostanziale continuità ed è interrotta solo per due brevi tratti in corrispondenza della valle del Sele e poco a nord di San Gregorio Magno. Queste interruzioni marcano in superficie i confini dei frammenti in cui è suddivisa la faglia principale (*si veda l'illustrazione nella pagina a fronte*). Nella fotografia è ripreso un tratto della scarpata che si trova in prossimità dell'epicentro; in questo settore essa raggiunge l'altezza di un metro. La scarpata è l'indizio più evidente delle deformazioni permanenti che il terremoto produce sulla superficie terrestre. Ben presto i processi erosivi smantelleranno il bordo della scarpata e accumuleranno sedimenti al suo piede, tendendo così a ripristinare l'originario equilibrio topografico. Dopo un tempo più o meno lungo, in relazione alle condizioni climatiche locali e alla natura dei terreni interessati, la scarpata verrà completamente ricoperta da depositi più recenti, ma le deformazioni prodotte dal terremoto saranno ancora ben conservate nel sottosuolo.

Gli studi paleosismologici si effettuano in genere aprendo trincee attraverso scarpate di faglia. Tali scavi hanno profondità e larghezza dell'ordine di tre-quattro metri e lunghezze da una decina ad alcune decine di metri. Nella fotografia è ripresa una delle trincee aperte lungo la scarpata di faglia prodotta dal terremoto del 23 novembre 1980, riconoscibile ancora oggi come una debole inflessione del terreno (*frecce*). La zona più vicina al fotografo appare ribassata rispetto a quella sullo sfondo. Sulla parete della trincea, in corrispondenza della scarpata del 1980, si osserva una chiara zona di deformazione in cui è riconoscibile un'inflessione simile a quella in superficie, ma più marcata. Questa zona ha registrato gli effetti di «antenati» del terremoto del 1980 che, come quest'ultimo, avevano prodotto una scarpata di faglia alta alcune decine di centimetri. L'alterazione della morfologia del Piano di Pecore conseguente alla formazione di una scarpata doveva influenzare profondamente la successiva sedimentazione, come suggerito dal diverso spessore degli orizzonti sui due lati della zona di deformazione (*si veda l'illustrazione alle pagine 20 e 21 in alto*).

ultimi che favoriscono la conservazione dei fenomeni tettonici di superficie, ha indirizzato in modo decisivo la ricerca verso la nuova disciplina.

Decine di faglie sono state già studiate in California, nel Nevada, nello Utah, consentendo di creare *ex novo* un catalogo sismico in regioni in cui la colonizzazione ha solo 150 anni di vita. Ma anche in regioni dotate di una storia più estesa, come nel bacino del Mediterraneo, la paleosismologia riveste un ruolo fondamentale nell'estendere all'indietro e rafforzare cataloghi sospetti di scarsa attendibilità o palesemente incompleti. Si consideri, infatti, che il tempo di ricorrenza caratteristico di numerose faglie è dell'ordine di molti secoli o di migliaia di anni, e che per motivi facilmente intuibili le dimensioni dei terremoti tendono a essere inversamente proporzionali alla loro frequenza. Ciò implica che alcuni cataloghi, spesso nella preoccupante inconsapevolezza di chi in seguito li usa per stime di rischio sismico, contengono solo occasionalmente i terremoti più grandi che possono aver luogo lungo le faglie della regione a cui si riferiscono, e vengono invece per lo più riempiti da eventi minori a breve periodo di ricorrenza.

Anche se oggi esistono vari metodi per l'identificazione e la datazione di grandi terremoti del passato, quello più diretto e intuitivo si basa sullo studio dei processi di alterazione della topografia che hanno luogo durante e a seguito di un grande terremoto. La deformazione legata allo scorrimento lungo la faglia sismogenetica interesserà i depositi sottostanti la superficie sotto forma di dislocazione, mentre i depositi in corso di sedimentazione registreranno l'evento come brusca alterazione dell'equilibrio del normale ciclo sedimentario di erosione-deposizione. Gli effetti dei processi innescati da questa condizione di instabilità costituiranno, assieme alla scarpata stessa, gli elementi della registrazione geologica del terremoto. In condizioni favorevoli tali effetti verranno preservati attraverso il seppellimento da parte di depositi più recenti, e potranno quindi essere messi in luce da appositi scavi a trincea le cui pareti «racconteranno» la storia geologica recente del sito.

La datazione di orizzonti ricchi in sostanze organiche posti nell'ambito di depositi di specifico significato sedimentologico consentirà di creare un quadro cronologico di riferimento per i paleoterremoti riconosciuti. Per esempio, l'età della parte alta dello strato sedimentario su cui si sono accumulati i detriti della scarpata, strato che costituiva la superficie topografica fino al verificarsi del terremoto, potrà intuitivamente essere assunta come di poco anteriore all'età del terremoto stesso. Analogamente, la datazione dei più bassi tra i depositi creati dal disfacimento della scarpata vincolerà un'età minima, permettendo quindi di circoscrivere il verificarsi del paleoterremoto entro un intervallo cronologico ristretto. Per gli scopi pratici, questa informazione si accompagnerà alle valutazioni sulle dimensioni del terremoto, che saranno state ricavate sulla base dell'estensione e dell'altezza della scarpata.

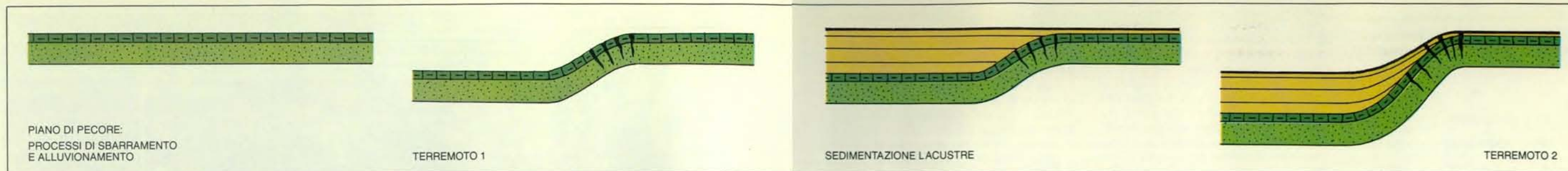
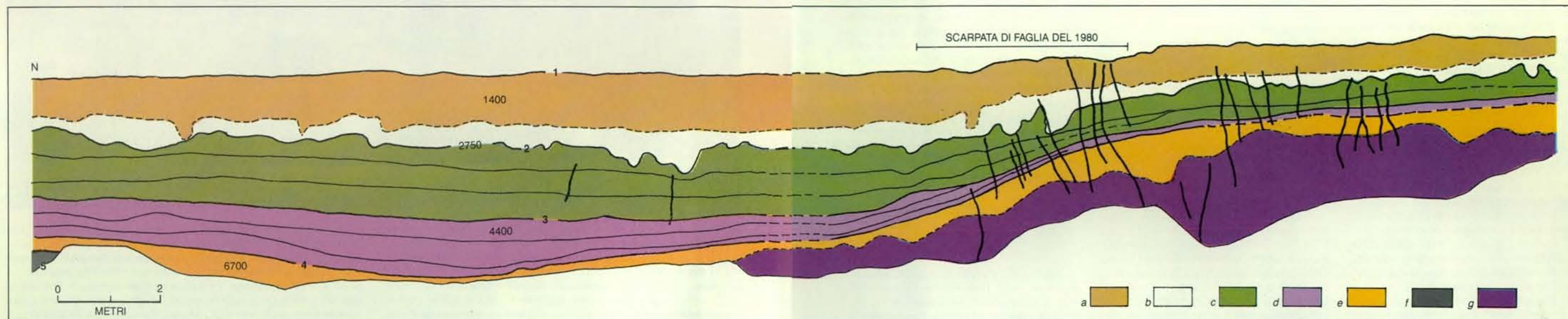
Tuttavia, anche se il ritrovamento e la datazione di chiare registrazioni geologiche di un paleoterremoto rappresentano già un risultato considerevole, l'informazione ottenuta non è ancora sufficiente a stabilire se esso sia o meno caratteristico per quella faglia, ovvero se possano verificarsene di più grandi, e a stimare il relativo tempo di ricorrenza. Si rende quindi opportuno individuare siti in cui sia possibile mettere in luce deformazioni legate ad almeno due paleoterremoti. La congruenza tra le deformazioni prodotte in superficie dai due eventi avvalorerà l'ipotesi che essi siano caratteristici, mentre il tempo intercorso tra di essi, o la media tra i diversi intervalli nel caso si siano osservati numerosi eventi, darà il tempo di ricorrenza cercato. Qualora questo risulti più breve del tempo già trascorso dall'ultimo terremoto ci troveremo di fronte a una tipica *lacuna sismica*; a questo punto il buon senso ci guiderà nel prendere le necessarie misure e precauzioni di carattere amministrativo, nonché nell'attivare ricerche geofisiche mirate al rilevamento di eventuali precursori sismici entro un'area geografica ormai ben delimitata.

Le trincee di Piano di Pecore hanno messo in luce una successione di depositi lacustri intercalati a depositi colluviali di ambiente subaereo provenienti dal disfacimento delle rocce che affiorano sui rilievi circostanti. I primi, contraddistinti da varie tonalità di grigio, sono costituiti prevalentemente da sabbie (*più scure*) e argille (*più chiare*) ben stratificate, di origine per lo più vulcanica. Questa origine non deve sorprendere, in quanto le cronache storiche riportano con frequenza la caduta di ceneri e sabbioni piroclastici in relazione ai principali episodi del vulcanismo vesuviano e flegreo. Questi depositi indicano, quindi, che almeno in alcune fasi della sua storia Piano di Pecore ospitava un lago, e che in esso confluivano trasportate dalle acque meteoriche le ceneri deposte sui rilievi circostanti. I livelli colluviali, riconoscibili per il colore marrone ocraceo, identificano invece periodi in cui il bacino, come avviene oggi, ospitava una palude il cui fondo veniva tappezzato da materiali vegetali in decomposizione. La datazione radiometrica di frammenti legnosi, più frequenti negli intervalli lacustri, o di livelli carboniosi, caratteristici degli intervalli palustri, ha consentito di inquadrare gli orizzonti incontrati nello scavo in un riferimento cronologico assoluto. Ciò ha indirettamente consentito di datarne l'epoca di deposizione e quindi, in ultima analisi, di datare i più recenti terremoti che hanno prodotto una scarpata di faglia a Piano di Pecore.

L'identificazione di faglie attive nell'Appennino meridionale

Nell'area mediterranea lo studio dei terremoti del passato e le valutazioni di pericolosità sismica sono stati fino a oggi principalmente basati sulla disponibilità di un cospicuo, e in larga misura unico, patrimonio di documenti e testimonianze storiche. L'analisi di questa enorme mole di informazioni, di cui i cataloghi rappresentano solo una sintesi, ha comportato la messa a punto di strumenti di indagine statistica estremamente complessi ed evoluti. Viceversa, le ricerche basate sull'osservazione di indizi di superficie del regime tettonico oggi attivo hanno segnato il passo, principalmente a causa delle difficoltà sorte nell'individuare tali indizi e nell'associarli al verificarsi di grandi terremoti storici.

La penisola italiana è nota agli studiosi di scienze della Terra di tutto il mondo per la complessità della sua struttura geologica. La conseguente difficoltà con cui la deformazione che ha luogo nello strato sismogenetico si propaga in superficie potrebbe rappresentare una spiegazione della scarsa evidenza delle faglie attive. Una ulteriore spiegazione potrebbe essere rappresentata dall'influenza degli agenti esogeni, che in un clima umido come quello delle zone interne della penisola possono obliterare rapidamente deboli modificazioni della superficie topografica legate ad attività tettonica. Sulla base di questi elementi si potrebbe giungere a ritenere che nella nostra penisola, e in generale in tutta l'area mediterranea, solo eventi catastrofici come quello del 1980 siano in grado di produrre effetti di superficie tali da consentire l'avvio di tutte le ricerche che su di essi sono basate.

Negli ultimi anni, tuttavia, è apparso chiaro che, anche se questa conclusione è sostanzialmente corretta, la sua giustificazione nei termini sopra espressi non lo è, ed è quindi necessario ripensare dalle radici tutto il problema. Il confronto tra la scarpata di faglia prodotta dal terremoto del 1980 e le scarpate descritte da Gilbert prima, e l'esperienza diretta di terreno effettuata sulle sue orme dopo, ci hanno finalmente indicato la stra-

La formazione di una scarpata di faglia altera notevolmente l'equilibrio morfologico, innescando processi di erosione e deposizione che tendono a ristabilire l'equilibrio di partenza. Le forme erosive e i depositi così prodotti costituiscono, con la scarpata di faglia, le «registrazioni geologiche» degli eventi sismici del passato. Lo studio delle trincee di Piano di Pecore ha mostrato come l'evoluzione del bacino possa essere il risultato della ripetizione del seguente ciclo: un primo terremoto provoca la formazione di una soglia nella zona di sbocco del drenaggio del Piano modificando la topografia, interrompendo il normale deflusso e determinando l'alluvionamento del bacino, che si trasforma in un lago. La sedimentazione lacustre porta gradualmente al colmamento della depressione e, quando lo spessore dei depositi raggiunge l'altezza della scarpata, viene a interessare anche una parte della zona di soglia fino alla restaurazione di una topografia pianeggiante e alla conseguente riapertura del drenaggio. Il terremoto successivo dà inizio a un nuovo ciclo.

Elaborando graficamente le osservazioni effettuate sulle pareti delle trincee è stato possibile riconoscere le deformazioni e i depositi associati a quattro terremoti precedenti quello del 1980 e a esso molto simili. Per ciascun evento è stata valutata l'entità della deformazione ed è stato delimitato un intervallo cronologico. Nell'illustrazione è rappresentata la sezione semplificata della trincea; diversi colori differenziano i depositi di origine lacustre (*c, d, f*) dai depositi colluviali (*a, b, e, g*), mentre a tratto spesso sono riportate faglie e fratture. I numeri in neretto identificano gli eventi sismici che possono essere «letti» nella trincea; il profilo in corrispondenza di ciascun numero rappresenta la traccia della superficie topografica al tempo del terremoto stesso. Tale superficie è stata deformata anche da tutti gli eventi successivi cosicché, se si confronta il dislivello esistente tra i due lati della faglia per le diverse superfici, si nota, per esempio, che esso risulta circa il doppio per la superficie 2 in rapporto alla 1. Il dislivello aumenta verso il basso, in accordo con la semplice osservazione che depositi più antichi hanno registrato più eventi sismici. Dalla trincea sono stati prelevati alcuni campioni di materiale carbonioso che sono stati datati con il metodo del carbonio 14. Per ogni campione è stata ottenuta un'età (riportata nel livello cui si riferisce) la cui precisione di misura è stimata in 50 anni. La datazione del primo orizzonte deformato da un evento e del primo non deformato immediatamente superiore consente di determinare i limiti cronologici entro cui l'evento ha avuto luogo.

da per risolvere questo dilemma. Infatti, per quale ragione sui versanti dell'Appennino non osserviamo quello che apparve a Gilbert durante le sue indagini sulla Sierra Nevada e sulle Montagne Rocciose? Perché la scarpata di faglia non si manifesta alla base del versante montuoso, come da lui descritto, ma lo taglia a metà, come è avvenuto nel corso del terremoto del 1980? Certamente gli sforzi tettonici cambiano nel tempo e, anche se un medesimo regime può permanere per milioni di anni, prima o poi si esaurirà cedendo il passo a un regime diverso. E così, mentre Gilbert descriveva qualcosa che agisce da almeno 10 milioni di anni e che è oggi nel pieno della sua maturità, noi assistiamo ai primi passi di un regime tettonico che sta subentrando a uno del tutto differente.

Naturalmente, il nuovo regime non sta modellando una superficie piatta, bensì una regione corrugata da immensi sforzi compressivi e oggi caratterizzata da profondi contrasti geologici e topografici. La scarpata che taglia in due un gruppo montuoso esemplifica e anticipa quello che sarà il panorama futuro dell'Appennino meridionale: dove erano valli e depressioni compariranno dorsali montuose, i vecchi rilievi scompariranno per l'azione combinata dell'erosione e della dislocazione tettonica, il reticolo idrografico verrà stravolto e ridisegnato. Ma tutto questo richiederà tempo e, mentre noi assisteremo al progressivo imporsi del nuovo regime tettonico, le faglie sismogenetiche resteranno subdolamente nascoste in mezzo alla moltitudine delle faglie ormai inattive, a meno che un terremoto come quello del 1980 non giunga a rivelarne la posizione.

Così descritto, il momento geologico che l'Appennino meridionale sta attraversando ci appare assolutamente unico, e la possibilità di testimoniarlo entusiasmante. Tuttavia, le implicazioni di questo stato di cose costituiscono motivo di seria preoccupazione per i sismologi. Come si diceva, le faglie che hanno generato terremoti disastrosi come quelli del 1456, del 1688, del 1694, del 1732, del 1857 non hanno una espressione superficiale ovvia; dunque semplicemente non sappiamo con certezza dove si trovano. Per di più, l'esperienza fatta con il terremoto del 1980 ci ha insegnato che, paradossalmente, è più facile che faglie chiaramente espresse alla superficie siano il prodotto di regimi tettonici ormai superati che non di quello oggi attivo. Sapere dove si trovano queste faglie per valutarne il potenziale è dunque compito arduo, e forse impossibile, se il terremoto del 1980 non ci avesse inaspettatamente fornito una chiave di lettura assolutamente eccezionale.

I predecessori del terremoto del 1980

Fino a questo punto, la constatazione della giovane età del regime tettonico in atto non è andata al di là della semplice osservazione qualitativa. Fatto ancora più importante, questa constatazione non ha offerto alcuna indicazione sulla velocità della deformazione tettonica, ovvero sulla velocità del rilascio sismico e quindi, in ultima analisi, sul tempo di ricorrenza dei grandi terremoti che interesseranno l'Appennino meridionale. Si è reso così essenziale elaborare un quadro cronologico di riferimento che serva a dare un senso pratico alle osservazioni finora condotte. Questo obiettivo è stato raggiunto attraverso un esperimento paleosismologico, il primo di questo tipo a essere condotto in Italia e in Europa, avviato da ricercatori dell'Istituto nazionale di geofisica di Roma lungo la scarpata prodotta dal terremoto del 1980.

Tra i numerosi siti che sembravano aver registrato deformazioni indotte da eventi precedenti a quello del 1980 è stato selezionato il Piano di Pecore di Colliano, una piccola depressione circolare posta a 1200 metri di quota lungo la dorsale del Monte Marzano-Monte Carpineta. Piano di Pecore, che si trova solo qualche chilometro a sud dell'epicentro strumentale del terremoto, fu vistosamente attraversato dalla scarpata di faglia prodotta dal terremoto del 1980. La posizione e la conformazione del bacino ne fanno una efficiente trappola per i sedimenti dilavati dai rilievi circostanti. Questo meccanismo è apparso particolarmente chiaro all'indomani del terremoto, quando la parte centrale del Piano si è trovata ribassata di circa 80 centimetri rispetto al suo unico punto di drenaggio. Il bacino ha così iniziato una fase di rapido colmamento che ha velocemente

EVENTO	ETÀ (ANNI)	RIGETTO VERTICALE (CENTIMETRI)
1	10	80
2	2750-3500	55
3	3500-4400	50
4	4400-6700	85
5	>6700	60

Nella tabella è riportato l'intervallo cronologico a cui è stato assegnato ciascuno dei quattro paleoterremoti riconosciuti nelle trincee e l'entità della dislocazione verticale prodotta da ognuno di essi. La somiglianza tra i valori di dislocazione porta a considerare il terremoto del 1980 come caratteristico per questa faglia. Sulla base di questi risultati si può calcolare un tasso di sollevamento medio di 0,4 millimetri all'anno e un tasso di estensione di 0,2 millimetri all'anno. Il tempo di ricorrenza medio per terremoti tipo 1980 risulta di circa 1700 anni.

sottratto agli agenti esogeni le deformazioni superficiali prodotte dal terremoto e che sta progressivamente ricreando le condizioni di drenaggio originarie.

L'ipotesi che un fenomeno simile si sia verificato anche a seguito di eventi del passato implica quindi che le deformazioni da questi prodotte si trovino oggi ben conservate nel sottosuolo di Piano di Pecore. Sono state così aperte due trincee che attraversano la scarpata al suo ingresso nel Piano e al centro di esso. L'esame di dettaglio delle pareti delle trincee ha consentito la ricostruzione di una sequenza stratigrafica costituita da una alternanza di depositi lacustri e colluviali (derivanti dal dilavamento dei versanti e deposti in ambiente subaereo) e che mostra, in corrispondenza della scarpata di faglia prodotta nel 1980, deformazioni legate ai diversi terremoti a esso precedenti. Attraverso un procedimento di ricostruzione grafica è stato possibile riconoscere, misurare e delimitare cronologicamente effetti deformativi legati a quattro paleoterremoti del tipo di quello del 1980.

I risultati provenienti dalle trincee di Piano di Pecore forniscono le prime misure assolute del tasso di deformazione tettonica cui è soggetto questo settore dell'Appennino. L'accumulo di un dislivello di 3,2 metri osservato tra i due lati della faglia in un intervallo di circa 7000 anni dà un tasso medio di sollevamento e di estensione in direzione NE-SO rispettivamente pari a 0,4 e 0,2 millimetri all'anno. Quest'ultima stima sale a quasi 0,4 millimetri all'anno ricordando che il terremoto del 1980 ha comportato fagliazione lungo una faglia parallela a quella attraverso cui sono state scavate le trincee e assumendo che questa modalità di rottura sia anch'essa comune a tutti gli eventi tipo 1980.

Attraverso la ragionevole assunzione che la meccanica del terremoto del 1980 sia simile a quella degli altri grandi terremoti che avvengono in regioni adiacenti della dorsale appenninica si rende possibile, almeno in prima approssimazione, estendere a queste ultime le stime ottenute a Piano di Pecore. Le stesse stime possono essere usate per quantificare attraverso un semplice calcolo l'età di inizio del regime tettonico oggi attivo. Le carte geologiche della zona riportano che nell'area circostante Piano di Pecore affiora una particolare formazione di calcari di età cretacica (da 140 a 65 milioni di anni fa) il cui spessore totale è valutato pari a 300 metri. Poiché essi si fronteggiano sui due lati della faglia, e tenendo opportunamente conto della loro pendenza rispetto a essa, si può concludere che la dislocazione verticale complessiva accumulata lungo la faglia deve essere pari o minore di 300 metri. Infine, sapendo che tale dislocazione aumenta alla velocità di 0,4 millimetri all'anno, si giunge a concludere che la faglia ha iniziato la sua attività non più di 750 000 anni fa, un'età molto breve rispetto a quella che presumibilmente sarà la durata di questa fase tettonica.

Il riconoscimento di quattro paleoeventi precedenti il terremoto del 1980 e a esso molto simili, almeno per quanto riguarda natura, geometria ed entità delle deformazioni osservate a Piano di Pecore, implica che questi eventi siano «caratteristici» nel senso sopra descritto. La tabella di questa pagina mostra che i tre eventi più antichi si sono succeduti con una cadenza piuttosto regolare, mentre quello immediatamente precedente il 1980 ha avuto luogo circa 1000 anni prima di quanto sarebbe stato lecito aspettarsi sulla base dei terremoti precedenti.

La cronologia dei paleoeventi succedutisi lungo la faglia sismogenetica che ha originato il terremoto del 1980 fa nascere un problema del tutto nuovo e decisamente preoccupante. Sappiamo infatti dalle fonti storiche che terremoti apparentemente simili a quello del 1980 si sono ripetuti con una cadenza di alcuni secoli. Uno in particolare, quello avvenuto nel 1694, sembra aver colpito nello stesso modo le stesse zone interessate dal terremoto del 1980, e infatti su questa base era stato fino a oggi comunemente ritenuto suo «gemello». Come si è visto, l'evidenza paleosismologica contrasta con questa ipotesi, ma allo stesso tempo non è ovviamente in grado di fornire una spiegazione alternativa.

Tra le ipotesi che è possibile avanzare vi è quella secondo cui il terremoto del 1694 sarebbe stato in effetti quel tanto più piccolo del suo successore da non superare la soglia energetica necessaria a produrre apprezzabili deformazioni della superficie, ma sufficientemente grande da provocare un quadro del danneggiamento apparentemente identico. E ancora, esso potrebbe avere avuto luogo lungo una struttura sismogenetica parallela e speculare a quella lungo cui è avvenuto il terremoto del 1980, così da mostrare una distribuzione del danneggiamento simile per movimento lungo una faglia che giunge in superficie a 15-20 chilometri da Piano di Pecore.

È molto probabile che la spiegazione di questo problema preveda una combinazione delle due ipotesi formulate. In ogni caso, il sospetto che gli effetti di questo come di molti altri terremoti storici appenninici siano stati decisamente sovrastimati, e quindi che alcune stime di rischio sismico basate sui cataloghi storici siano inutilmente pessimistiche, resta. L'apertura di nuove trincee in nuovi punti della faglia, la progressiva chiarificazione dei rapporti tra questa e le faglie adiacenti e una auspicabile rilettura critica della documentazione su particolari eventi storici dovrebbero consentirci nei prossimi anni di compiere significativi progressi nella valutazione del potenziale sismogenetico della dorsale appenninica e nell'individuazione delle aree più immediatamente minacciate dal verificarsi di grandi terremoti come quello del 1980.

BIBLIOGRAFIA

SCHWARTZ DAVID P. e COPPERSMITH KEVIN J., *Fault Behavior and Characteristic Earthquakes: Examples from the Wasatch and San Andreas Fault Zones* in «Journal of Geophysical Researches», 89, pp. 5681-5698, 1984.

WESTAWAY ROBERT e JACKSON JAMES, *The Earthquake of 1980 November 23 in Campania-Basilicata (Southern Italy)* in «Geophysical Journal of the Royal Astronomical Society», 90, pp. 375-443, 1987.

BOSCHI ENZO e GUIDOBONI EMANUELA, *I grandi terremoti medievali in Italia* in «Le Scienze» n. 249, maggio 1989.

PANTOSTI DANIELA e VALENSISE GIANLUCA, *Riconoscere il «terremoto caratteristico»: il caso dell'Appennino centro-meridionale* in *I terremoti prima del Mille* a cura di Emanuela Guidoboni, Storia Geofisica Ambiente Editore, Bologna, 1989.

BERNARD PASCAL e ZOLLO ALDO, *The Irpinia (Italy) 1980 Earthquake: Detailed Analysis of a Complex Normal Fault* in «Journal of Geophysical Researches», 94, pp. 1631-1648, 1989.

PANTOSTI DANIELA e VALENSISE GIANLUCA, *Faulting Mechanism and Complexity of the 23 November 1980, Campania-Lucania Earthquake, Inferred from Surface Observations* in «Journal of Geophysical Researches», 95, pp. 15 319-15 341, 1990.

Oltre il superconduttore 1-2-3

Nuovi orizzonti si stanno aprendo nel campo dei superconduttori ad alta temperatura, compresi quelli contenenti piani di atomi di rame e ossigeno che offrono agli elettroni vere e proprie «autostrade» su cui muoversi

di Robert J. Cava

Più di dieci anni fa, all'Università di Caen, in Francia, Bernard Raveau e collaboratori cominciarono a sintetizzare una serie di composti costituiti da lantanio, bario, rame e ossigeno. Se avessero raffreddato qualcuno di questi materiali ceramici a temperature prossime a 40 kelvin, i ricercatori avrebbero potuto osservare che essi non offrivano alcuna resistenza al passaggio della corrente elettrica. Allora tuttavia ben pochi fisici, chimici o esperti di materiali avevano motivo di ritenere che quei composti fossero superconduttori e tanto meno erano disposti a credere alla possibilità di realizzare in pratica superconduttori con una temperatura di transizione così elevata. I materiali ceramici vennero relativamente trascurati dai ricercatori per parecchi anni, fino a quando, nel 1986, i fisici K. Alex Müller e J. Georg Bednorz dei laboratori di ricerca della IBM presso Zurigo si resero conto che essi costituivano la chiave per produrre una nuova famiglia di superconduttori ad alta temperatura capaci di dare nuovo impulso alla teoria e alle applicazioni della superconduttività. La scoperta, per la quale i due ricercatori hanno ricevuto il premio Nobel, ha dato l'avvio a una delle più straordinarie rivoluzioni della fisica dello stato solido.

Spronati dalla speranza di scoprire il *non plus ultra* dei superconduttori ad alta temperatura, nel 1987 quasi tutti coloro che avevano a disposizione un sistema periodico degli elementi e un forno si misero d'impegno a fabbricare i composti più bizzarri a base di rame e ossigeno. Tuttavia la natura è molto esigente per quanto riguarda la formazione degli ossidi di rame. Benché nella scelta degli elementi da combinare per ottenere i nuovi composti ci si possa far guidare dalle teorie della chimica dello stato solido, questi materiali sono così complessi che nessuna teoria è ancora in grado di fornire previsioni attendibili sulla loro struttura o sul loro comportamento. Bisogna sperare che intuito, preparazione teorica e fortuna portino proprio a quella giusta combinazione di elementi che consenta di ottenere una temperatura di transizione ancora più alta.

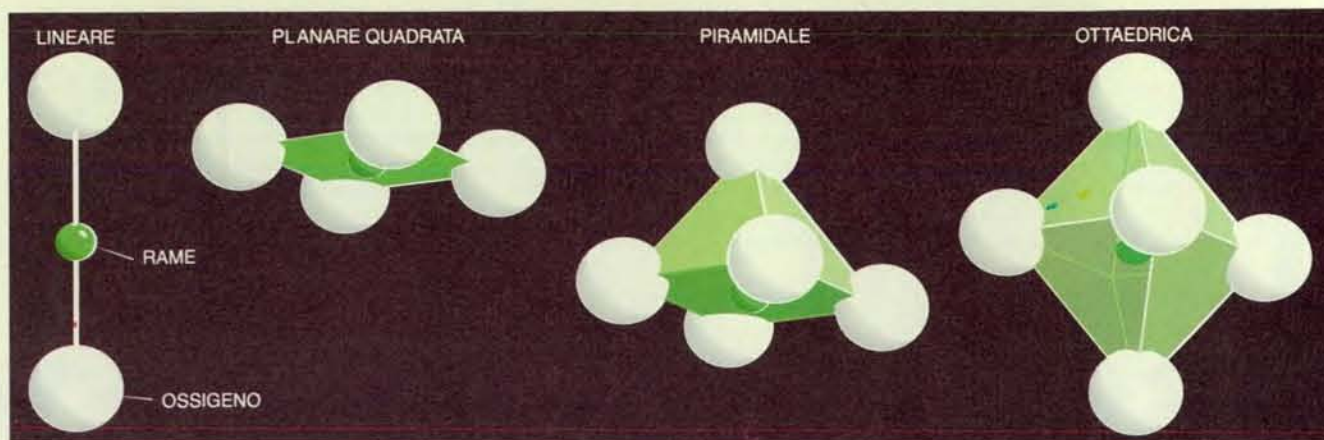
Nel febbraio 1987, Ching-Wu (Paul) Chu dell'Università di Houston, Maw-Kuen Wu dell'Università dell'Alabama e collaboratori sostituirono, nel composto sintetizzato da Bednorz e Müller, il lantanio con un'altra terra rara dall'atomo più piccolo, l'ittrio, e ottennero un superconduttore con una temperatura di transizione di 90 kelvin. Poco dopo R. Bruce van Dover, Bertram Batlogg e io, lavorando presso gli AT&T Bell Laboratories, determinammo per primi la composizione chimica del superconduttore, oggi noto con il nome 1-2-3 che esprime il rapporto tra gli atomi di ittrio, di bario e di rame. L'1-2-3 era il primo superconduttore con una temperatura di transizione superiore a 77 kelvin e poteva quindi essere raffreddato facilmente e con poca spesa in un bagno di azoto liquido. I migliori superconduttori metallici tradizionali, come le leghe di niobio e stagno, richiedevano come refrigerante l'elio liquido, che è più costoso, e la stessa cosa valeva per il primo superconduttore ad alta temperatura scoperto da Bednorz e Müller (si veda l'articolo *Le future applicazioni dei nuovi superconduttori* di Alan M. Wolsky, Robert F. Giese ed Edward J. Daniels in «Le Scienze» n. 248, aprile 1989).

Negli ultimi quattro anni l'impegno di migliaia di ricercatori in tutto il mondo ha permesso di produrre una decina di superconduttori con una temperatura di transizione superiore a 40 kelvin e alcuni con una temperatura di transizione superiore a 77 kelvin. Quasi tutti i superconduttori ad alta temperatura sono stati scoperti da sperimentatori che seguivano un'altra linea di ricerca e che, superati per ingegno (e aiutati) dalla natura, hanno trovato per pura coincidenza qualcosa di diverso. La più alta temperatura di transizione finora raggiunta è quella sbalorditiva di 125 kelvin relativa a un composto di tallio, bario, calcio, rame e ossigeno.

Per quanto gli studiosi non conoscano i limiti della superconduttività e neppure comprendano le interazioni fondamentali grazie alle quali essa si manifesta nei superconduttori ceramici, è possibile indicare le caratteristiche molecolari che favoriscono od ostacolano la superconduttività. L'aspetto più importante è che i reticoli cristallini di tutti i superconduttori ad alta temperatura contengono piani di atomi di rame e di ossigeno intercalati a strati di altri elementi. Quando un superconduttore ad alta temperatura viene portato a una temperatura inferiore a quella di transizione, i piani di atomi di rame e ossigeno offrono agli elettroni vere e proprie autostrade su cui spostarsi. Gli altri elementi nel reticolo cristallino possono essere scelti e disposti in modo tale da far aumentare o diminuire la temperatura di transizione al regime di superconduzione.

Un superconduttore costituito da tallio, bario, calcio, rame e ossigeno, di formula $Tl_2Ba_2Ca_2Cu_3O_{10}$, ha la più alta temperatura di transizione finora nota: quando viene raffreddato sotto i 125 kelvin non oppone alcuna resistenza al passaggio di corrente. Come altri superconduttori ceramici, questo composto del tallio contiene piani di atomi di rame e di ossigeno che permettono agli elettroni di spostarsi nel materiale.

I materiali scoperti recentemente da Bednorz, Müller e altri si comportano, sotto un certo aspetto, come normali superconduttori metallici. Se si applica una tensione a uno di questi materiali a temperatura ambiente, gli elettroni cominciano a muoversi al suo interno. Poiché vengono diffusi dagli atomi del composto gli elettroni perdono energia e questo fenomeno si manifesta nella resistenza elettrica. Se il materiale viene raffreddato, la resistenza e la perdita di energia decrescono. Quando si scende appena un po' al di sotto della temperatura di transizione, d'un tratto la resistenza si annulla. Gli elettroni subiscono una transizione di fase e non possono più essere diffusi dagli atomi.

I composti di Bednorz e Müller appartengono a un gruppo piuttosto ristretto di materiali ceramici con buone proprietà di conduzione. (È ben noto invece che le ceramiche comuni, quelle con cui si fanno i piatti e i lavandini, sono isolanti.) Ciascuno dei metalli presenti - lantanio, bario e rame - reagisce facilmente con l'ossigeno, formando il rispettivo ossido. Questi ossidi sono isolanti, cioè conducono male l'elettricità a temperatura ambiente. Tuttavia, combinando nelle giu-

La caratteristica elettronica fondamentale di quasi tutti i superconduttori ad alta temperatura è il legame fra rame e ossigeno. Un atomo di rame si può combinare con l'ossigeno secondo quattro strutture diverse che possono essere rappresentate tramite un poliedro di coordinazione, cioè una figura geometrica che comprende l'atomo di rame e gli atomi di ossigeno a cui esso è direttamente legato.

ste proporzioni lantanio, bario, rame e ossigeno, Bednorz e Müller ottennero un materiale ceramico che conduce bene l'elettricità a temperatura ambiente ed è superconduttore a 28 kelvin.

La natura isolante oppure conduttrice di un materiale ceramico dipende da come si comportano in seno al composto i portatori di carica negativa, cioè gli elettroni. L'ossido di bario (BaO), per esempio, è un isolante perché gli elettroni sono confinati nella regione circostante i nuclei degli atomi di bario e di ossigeno. Gli elettroni sono localizzati a causa dei vincoli di carica e di energia cui sono soggetti il bario e l'ossigeno. Il nucleo del bario contiene 56 protoni e quindi nello stato atomico neutro è circondato da 56 elettroni. Il nucleo dell'ossigeno ha otto protoni e quindi nello stato atomico neutro è circondato da otto elettroni. Per formare composti chimici come l'ossido di bario, gli elettroni si ridistribuiscono tra gli atomi di bario e di ossigeno in modo da realizzare lo stato con l'energia più bassa possibile.

Gli elettroni si distribuiscono intorno ai nuclei in orbitali, che si possono immaginare come gusci concentrici contenenti un certo numero di elettroni. Gli elettroni dei gusci interni sono fortemente legati al nucleo, mentre quelli dei gusci esterni si possono estrarre più facilmente e hanno un ruolo più importante nei fenomeni chimici ed elettrici.

L'ossigeno è il «collante» che tiene insieme i composti ceramici. Un atomo di ossigeno ha sei elettroni nel suo guscio esterno, che ne può contenere fino a otto. Dato che un guscio esterno completo (ottetto) è una configurazione molto più stabile rispetto a uno parzialmente occupato, acquisendo due elettroni un atomo di ossigeno può raggiungere uno stato di energia molto inferiore. Quindi un atomo di ossigeno «preferisce» avere una carica negativa pari a due: in altre parole, ha numero di ossidazione −2. Un atomo di bario ha due elettroni nel guscio esterno, ma perdendoli può portarsi in uno stato di energia molto più bassa: dunque il suo numero di ossidazione è +2. Per ottenere la configurazione elettronica di energia minima nell'ossido di bario ciascun atomo di bario cede due elettroni a un atomo di ossigeno. Dato che ci vuole molta energia per aggiungere o sottrarre un elettrone all'orbitale completo, ora gli elettroni sono fortemente localizzati e non sono in grado di trasportare corrente elettrica.

Questo scambio di elettroni costituisce anche la base del legame che tiene insieme il materiale ceramico. Gli atomi di bario con carica positiva (+2) attraggono quelli di ossigeno che hanno carica negativa (−2): questo tipo di legame si chiama ionico.

Virtualmente in tutti i materiali ceramici gli elettroni esterni degli atomi del metallo si trovano in uno stato di energia molto più alta rispetto agli elettroni esterni degli atomi di ossigeno. Quindi per raggiungere una configurazione stabile a bassa energia gli atomi del metallo di solito cedono uno o più elettroni dei gusci esterni, che vengono catturati dagli atomi di ossigeno. Poiché gli elettroni sono localizzati presso gli atomi del metallo e quelli di ossigeno, il materiale ceramico si comporta da isolante.

Molti materiali ceramici contenenti ossido di rame conducono bene l'elettricità perché alcuni degli elettroni sono liberi di passare da un atomo all'altro. In questo tipo di materiale ceramico gli elettroni non sono localizzati, a causa di una particolare interazione tra rame e ossigeno. In composti come Cu_2O, il rame cede facilmente un elettrone all'ossigeno e rimane quindi con il guscio esterno completo. I due atomi di rame (ciascuno con numero di ossidazione +1) formano un legame ionico con l'ossigeno (numero di ossidazione −2) e, dato che gli elettroni sono localizzati presso i nuclei, il materiale si comporta da isolante.

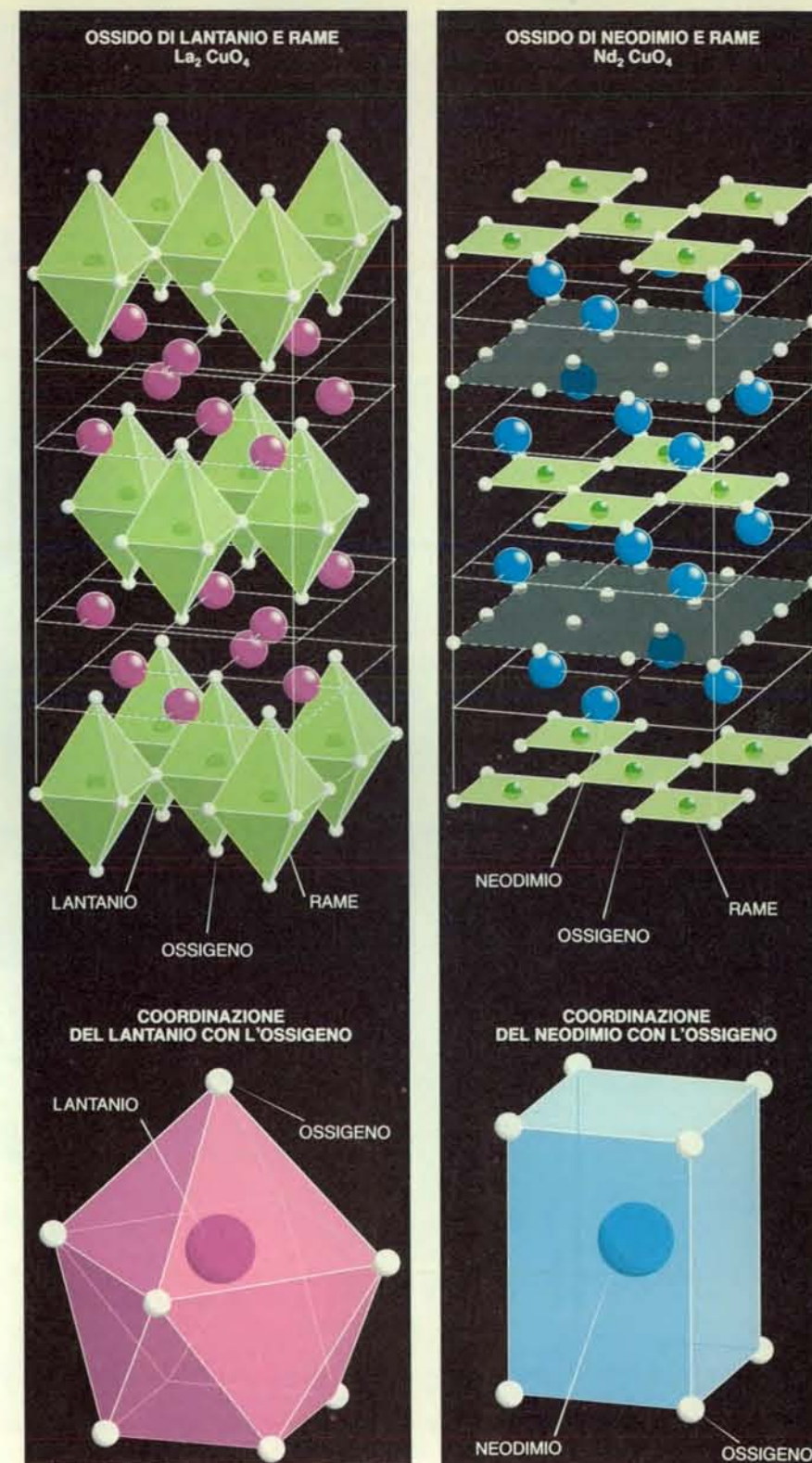
Invece in ossidi come CuO il rame non cede il secondo elettrone con la stessa facilità del primo, perché la perdita del secondo elettrone crea una lacuna nel guscio esterno. L'ossigeno deve «lottare» per ottenere il secondo elettrone e completare così il proprio orbitale esterno. Alla fine l'ossigeno «vince» perché la configurazione con un orbitale esterno completo nell'ossigeno è un poco più stabile di quella con un orbitale esterno completo nel rame. Tuttavia se nel reticolo cristallino sono presenti altri elementi oltre al rame e all'ossigeno, questo delicato equilibrio energetico può essere alterato e rame e ossigeno mettono in comune gli elettroni, in modo da completare i propri gusci esterni. (Questi elettroni condivisi danno luogo a un legame covalente fra ossigeno e rame.) Dato che gli elettroni sono liberi di muoversi fra gli atomi di rame e quelli di ossigeno, i materiali che contengono rame, ossigeno e altri elementi possono risultare buoni conduttori.

Nel legame chimico con l'ossigeno il rame può cedere uno, due o tre elettroni. Fra questi tre stati di ossidazione, il più stabile è +2; perciò lo stato +1 (che si trova per esempio in Cu_2O) si chiama rame «ridotto» e lo stato +3 (che si trova per esempio nell'ossido di sodio e rame, $NaCuO_2$) è detto rame «ossidato».

Nei superconduttori ceramici il rame sembra comportarsi come se avesse un numero di ossidazione frazionario. Il numero di ossidazione del rame dipende dall'influenza esercitata dall'ossigeno e dagli altri atomi del superconduttore: se è uguale a +2, gli elettroni sono localizzati nel legame rame-ossigeno. In certi casi, quando nel composto sono presenti altri elementi ossidanti, come lantanio e bario, da alcuni atomi di rame vengono estratti più di due elettroni, e il numero di ossidazione si sposta verso +3. In altri casi, elementi «riducenti» possono fornire elettroni ad alcuni atomi di rame, portando il numero di ossidazione da +2 verso +1. In entrambi i casi gli elettroni non sono più localizzati e possono partecipare alla conduzione elettrica.

Benché questa sorta di computo del numero di ossidazione sia essenziale per capire i superconduttori ceramici, si tratta di una semplificazione. In questi superconduttori gli atomi di ossigeno e di rame condividono alcuni elettroni per soddisfare i propri vincoli energetici. Gli elettroni condivisi si spostano dagli atomi di rame a quelli di ossigeno, formando un «mare» di elettroni o, più esattamente, una banda di conduzione. Se gli atomi di rame sono ridotti a numeri di ossidazione inferiori a +2, cioè se cedono in media meno di due elettroni, allora gli elettroni che si spostano nella banda di conduzione sono pochi. Se gli atomi di rame sono ossidati a numeri di ossidazione maggiori di +2, cioè se cedono in media più di due elettroni, allora la banda di conduzione brulica di elettroni e si generano regioni di carica positiva dette buche. Il concetto di numero di ossidazione frazionario ha senso solo perché in realtà si ha un'aggiunta o una sottrazione di elettroni a una banda di conduzione.

Il rame non è l'unico metallo che può assumere numeri di ossidazione frazionari in presenza di ossigeno. Il bismuto e il piombo hanno, come il rame, orbitali elettronici con vincoli energetici analoghi a quelli dell'ossigeno. In condizioni opportune entrambi i metalli formano legami covalenti con l'ossigeno, il che consente agli elettroni di spostarsi liberamente all'interno di una banda di conduzione. Sono stati scoperti ossidi sia di piombo sia di bismuto che sono superconduttori a temperature relativamente elevate; si discute ancora animatamente se questi composti siano superconduttori per lo stesso motivo per cui lo sono gli ossidi di rame.

La famiglia di superconduttori scoperta da Bednorz e Müller si basa sulla modificazione chimica dell'ossido di lantanio e rame (La_2CuO_4), la cui struttura cristallina è mostrata nell'illustrazione in questa pagina. Le strutture dei superconduttori ceramici sono di solito descritte in termini di poliedri di coordinazione, uno dei concetti chiave della chimica dello stato solido. Un poliedro di coordinazione è una figura geometrica che comprende un atomo metallico e gli atomi di ossigeno più vicini, ai quali esso è legato direttamente. I poliedri mettono in evidenza quanti atomi di ossigeno uno ione «preferisce» nel suo intorno immediato di legame. Questo numero preferito di atomi d'ossigeno può variare, per esempio, da due per il rame +1 a 15 per il cesio +1. Il numero effettivo di atomi di ossigeno con i quali uno ione è coordinato dipende sia dalla dimensione dello ione, sia dalle particolari esigenze degli altri atomi di metallo dell'ossido.

La struttura cristallina dell'ossido di rame e lantanio, La_2CuO_4 (*in alto a sinistra*), è alla base di molti superconduttori. Se alcuni atomi di lantanio sono sostituiti da atomi di calcio, di stronzio o di bario si ottiene un superconduttore. In basso a sinistra è illustrato il poliedro di coordinazione del lantanio con l'ossigeno. I superconduttori derivati da La_2CuO_4 sono di tipo *p* perché contengono portatori di carica positiva detti buche. Superconduttori di tipo *n* si possono ottenere dall'ossido di rame e neodimio, Nd_2CuO_4 (*in alto a destra*), sostituendo alcuni atomi di neodimio con atomi di cerio o di torio. Nei superconduttori derivati da Nd_2CuO_4 i portatori di carica negativa sono elettroni. Un atomo di neodimio è coordinato con otto atomi di ossigeno in un prisma a base quadrata (*in basso a destra*).

In La_2CuO_4 gli atomi di rame sono coordinati con sei atomi di ossigeno posti ai vertici di un ottaedro allungato. Per i vincoli energetici cui è sottoposto il guscio esterno del rame nello stato di ossidazione +2, gli atomi di ossigeno situati su due vertici opposti dell'ottaedro sono sempre più lontani dal rame che non gli atomi di ossigeno situati sugli altri quattro vertici. Fu questa distorsione strutturale, detta effetto Jahn-Teller, a indurre Bednorz e Müller a considerare gli ossidi di rame come potenziali superconduttori. L'effetto fa prevedere una forte interazione fra gli elettroni e le posizioni occupate dagli atomi di rame e ossigeno nel reticolo cristallino, e questa interazione è considerata molto importante per il passaggio al regime di superconduzione.

In La_2CuO_4 gli ottaedri di rame e ossigeno condividono i vertici occupati dai quattro atomi di ossigeno più vicini al rame. Gli atomi di rame e gli atomi di ossigeno a essi più vicini giacciono tutti su uno stesso piano. Questo piano definito da rame e ossigeno è la caratteristica elettronica fondamentale di tutti gli ossidi di rame superconduttori conosciuti. In questa zona della struttura cristallina hanno origine i portatori di carica che determinano la superconduttività.

La struttura cristallina di La_2CuO_4 è notevole perché i poliedri di coordinazione del lantanio e del rame si combinano in modo da riempire uno spazio tridimensionale con strutture bidimensionali. Pertanto i piani di rame e ossigeno sono intercalati a un doppio strato inerte di lantanio e ossigeno.

L'ossido di lantanio e di rame non è superconduttore a causa di un'interazione che si verifica fra gli elettroni dei gusci esterni di atomi di rame vicini. Un atomo di rame cede all'ossigeno due elettroni e ne conserva nove nel guscio esterno. Ciascuno dei nove elettroni ha un momento magnetico, cioè si comporta come se fosse un minuscolo magnete a barra con un polo nord e un polo sud. L'energia derivante da questa interazione magnetica diventa minima quando otto di questi nove elettroni si accoppiano: i loro momenti magnetici si elidono assumendo direzioni parallele e versi opposti. Quando i restanti elettroni non accoppiati su atomi di rame vicini interagiscono tra loro e si allineano, i loro momenti magnetici puntano in versi opposti. Questo fenomeno, detto antiferromagnetismo, fissa gli elettroni al reticolo cristallino e impedisce non solo la superconduttività, ma addirittura la normale conduttività.

La superconduttività si manifesta solamente se l'antiferromagnetismo viene completamente eliminato. Per questa ragione molti studiosi ritengono che l'origine del magnetismo sia intimamente legata a quella della superconduttività ad alta temperatura. Alcuni hanno attribuito a questo legame una connotazione sinistra, definendo magnetismo e superconduttività i «Jekyll e Hyde» della superconduttività ad alta temperatura.

La struttura cristallina di La_2CuO_4 può essere però modificata per eliminare l'antiferromagnetismo e introdurre la superconduttività. Bednorz e Müller ottennero il loro famoso superconduttore ad alta temperatura sostituendo alcuni degli atomi di lantanio con atomi di bario. La formula chimica risultante è $La_{2-x}Ba_xCuO_4$.

In questo composto si conserva la neutralità di carica se per ogni atomo di lantanio sostituito da uno di bario un atomo di rame si ossida da +2 a +3. L'elettrone in più ceduto dal rame non è localizzato e va nella banda di conduzione. Quando gli atomi di rame raggiungono un numero di ossidazione critico, prossimo a +2,2, l'antiferromagnetismo scompare e compare la superconduttività.

In $La_{2-x}Ba_xCuO_4$ il rame è coordinato con sei atomi di ossigeno, che formano i vertici di un ottaedro allungato. Poiché il numero di coordinazione cresce grosso modo con le dimensioni dell'atomo, gli atomi di lantanio, che sono più grandi degli atomi di rame, sono coordinati con nove atomi di ossigeno. Anche gli atomi di bario, la cui grandezza è prossima a quella del lantanio, sono coordinati con nove atomi di ossigeno. In seno al reticolo cristallino gli atomi di bario possono dunque occupare le stesse posizioni degli atomi di lantanio e, in effetti, vi si distribuiscono a caso. Una siffatta distribuzione di atomi in posizioni cristallografiche fisse si chiama soluzione solida. (Una soluzione liquida consiste invece in una distribuzione casuale di ioni in posizioni casuali.)

Il composto 1-2-3 ($YBa_2Cu_3O_7$) è solamente uno dei termini di quella che i chimici chiamano serie omologa. Ciascun termine della serie si può ricavare da un altro per aggiunta o sottrazione di semplici componenti strutturali. Tre termini della serie omologa - $YBa_2Cu_3O_7$, $Y_2Ba_4Cu_7O_{15}$ e $YBa_2Cu_4O_8$ - sono superconduttori con una temperatura di transizione superiore a 77 kelvin. Non si sa ancora se i composti $YBaCu_2O_5$ (che qui non è illustrato) e $Y_2Ba_3Cu_5O_{12}$ esistano. Si prevede che le caratteristiche strutturali di $YBaCu_2O_5$ siano simili a quelle di $YBaCuFeO_5$, che non è un ossido superconduttore.

Il concetto di soluzione solida è fondamentale per il manifestarsi della superconduttività negli ossidi di rame. Di regola le soluzioni solide si formano solo se le grandezze degli ioni sono circa uguali e se i legami chimici che si costituiscono sono di tipo simile. Quindi il bario, lo stronzio o il calcio possono formare soluzioni solide in La_2CuO_4 sostituendosi in parte al lantanio. Quando in un cristallo di La_2CuO_4 uno su nove atomi di lantanio è sostituito dallo stronzio, si ha il composto $La_{1,8}Sr_{0,2}CuO_4$, dove il numero di ossidazione del rame è +2,2. In questo composto la temperatura di transizione alla superconduttività è di 40 kelvin, la più elevata per le soluzioni solide basate su La_2CuO_4.

A differenza delle soluzioni solide che rendono superconduttore La_2CuO_4, il composto $YBa_2Cu_3O_7$ (il superconduttore 1-2-3) ha un reticolo cristallino ordinatissimo, costituito da piani paralleli di ioni di ittrio, bario e rame (*si veda l'illustrazione in queste due pagine*). Gli atomi degli elementi metallici non scambiano le proprie posizioni nel reticolo cristallino. I piccoli ioni di ittrio +3 sono sempre legati a otto atomi di ossigeno, i grandi ioni di bario +2 a dieci atomi di ossigeno. Gli atomi di rame situati fra un piano di atomi di bario e uno di atomi di ittrio sono coordinati con l'ossigeno in strutture piramidali. Le basi di queste piramidi di rame e ossigeno si fronteggiano reciprocamente ai due lati di un piano di atomi di ittrio e costituiscono i piani di atomi di rame e ossigeno necessari alla superconduttività. Gli atomi di rame situati tra due strati consecutivi di bario sono coordinati con quattro atomi di ossigeno e formano rombi (coordinazione planare quadrata), perpendicolari ai piani occupati dal bario, i cui vertici sono collegati a costituire una catena.

Le particolarità del legame rame-ossigeno fanno sì che uno spazio tridimensionale venga riempito da una combinazione di catene unidimensionali e di basi bidimensionali delle piramidi. Una volta uno scienziato egiziano mi disse scherzosamente che i suoi compatrioti dovrebbero essere molto abili nella ricerca di nuovi superconduttori, dato che la magia delle piramidi è alla radice della loro tradizione culturale.

In $YBa_2Cu_3O_7$, il rame ha un numero di ossidazione medio di +2,33 (sette dei 14 elettroni richiesti dall'ossigeno sono forniti da un atomo di ittrio e da due di

bario; gli altri sette elettroni sono ceduti da tre atomi di rame). Si è dimostrato che tanto nelle strutture a catena quanto nelle basi delle piramidi gli atomi di rame hanno più o meno lo stesso numero di ossidazione.

Nel composto $YBa_2Cu_3O_7$, i sette atomi di ossigeno sono d'importanza cruciale per la superconduttività. Se il numero di atomi di ossigeno viene ridotto da sette a sei, si forma $YBa_2Cu_3O_6$, che è isolante. L'ossigeno viene rimosso da una sola posizione cristallografica e le strutture planari quadrate di rame e ossigeno si trasformano in strutture lineari; queste ultime costituiscono una geometria di coordinazione stabile per il rame nello stato di ossidazione +1. La diminuzione del contenuto di ossigeno non influisce direttamente sulle strutture piramidali di rame e ossigeno, ma fa sì che gli atomi di rame in queste ultime abbiano ora numero di ossidazione +2. In questa configurazione gli elettroni sono localizzati per effetto dell'antiferromagnetismo.

Quando il contenuto di ossigeno di $YBa_2Cu_3O_6$ viene aumentato, i nuovi atomi si collocano direttamente nell'intorno di legame delle strutture lineari; perciò il rame si ossida e passa dal numero di ossidazione +1 a +2. L'ossigeno non forma legami a caso con le strutture lineari, ma tende a costituire il massimo numero possibile di strutture planari quadrate di rame e ossigeno.

Quando il contenuto di ossigeno raggiunge un livello tale per cui vi sono in media 6,5 atomi di ossigeno disponibili per tre atomi di rame, l'ossigeno nelle strutture piramidali sottrae al rame un numero di elettroni sufficiente a indurre la superconduttività. Il composto risultante, $YBa_2Cu_3O_{6,5}$, ha una temperatura di transizione di 60 kelvin. Questo composto ha una struttura cristallina ben definita, in cui catene di strutture planari quadrate e lineari formano una schiera ordinata. Via via che il contenuto di ossigeno aumenta, continuano a formarsi nuove strutture planari quadrate finché, quando le catene sono completate, la temperatura di transizione al regime di superconduzione balza a 90 kelvin.

Gli atomi di rame, sia in $YBa_2Cu_3O_7$ sia in $La_{1,8}Ba_{0,2}CuO_4$, hanno numeri di ossidazione maggiori di +2. Di conseguenza nella banda di conduzione vengono introdotte buche cariche positivamente; questi materiali sono chiamati dunque superconduttori di tipo *p*. I superconduttori in cui i portatori di carica sono elettroni sono detti di tipo *n*. Fino al 1988 tutti i superconduttori conosciuti erano di tipo *p*, tanto che molti ritenevano che non si sarebbe mai ottenuto un superconduttore di tipo *n* a base di ossido di rame.

Invece nel 1988 Y. Tokura, H. Takagi e Shin-ichi Uchida dell'Università di Tokyo scoprirono il primo, e finora unico, superconduttore ceramico di tipo *n*. Il nuovo superconduttore era basato sull'ossido di rame e neodimio (Nd_2CuO_4), la cui struttura è simile a quella di La_2CuO_4. Ma mentre il lantanio preferisce intorno a sé nove atomi di ossigeno, lo ione neodimio, più piccolo, è di solito coordinato con otto atomi di ossigeno. Nel composto Nd_2CuO_4 gli atomi di ossigeno circondano gli atomi di neodimio formando un prisma a base quadrata. Di conseguenza gli atomi di rame sono coordinati con quattro atomi di ossigeno in strutture planari quadrate parallele ai piani occupati dal neodimio.

Nei superconduttori a base di ossido di rame e neodimio alcuni atomi di neodimio sono sostituiti dal cerio o dal torio, con formazione di una soluzione solida. Tanto il cerio quanto il torio hanno numero di ossidazione +4 e dimensioni adatte per sostituire alcuni degli ioni di neodimio +3. Pertanto possono formarsi le soluzioni solide $Nd_{2-x}Ce_xCuO_4$ e $Nd_{2-x}Th_xCuO_4$. Quando x è uguale a

I superconduttori ad alta temperatura, come questo composto di piombo, stronzio, disprosio, calcio, rame e ossigeno di formula $Pb_2Sr_2Dy_{1-x}Ca_xCu_3O_8$ formano in genere cristalli rettangolari. I cristalli qui illustrati hanno uno spessore di circa 100 micrometri.

COMPOSIZIONE BASE DEGLI OSSIDI DI RAME SUPERCONDUTTORI E ALCUNE VARIANTI

STRUTTURA DI PARTENZA	SOSTITUZIONI	TEMPERATURA DI TRANSIZIONE (KELVIN)	COMMENTI
$La_{2-x}Ba_xCuO_4$	Ca o Sr	20-40	IL PRIMO SUPERCONDUTTORE A BASE DI OSSIDO DI RAME FU UN MATERIALE CERAMICO DI La, Ba, Cu E O
$Nd_{2-x}Ce_xCuO_{4-y}F_y$	Pr, Sm o Eu Th	10-25	QUESTA FAMIGLIA COMPRENDE I SOLI SUPERCONDUTTORI CERAMICI DI TIPO *n*
$La_{1,8-x}Sm_xSr_{0,2}CuO_4$	Eu, Gd, Tb o Dy	20	
$YBa_2Cu_3O_7$	La, Nd, Sm, Eu, Gd, Dy, Ho, Er, Tm, Yb o Lu Cu_4O_8 o Cu_7O_{15}	80-93	$YBa_2Cu_3O_7$ FU IL PRIMO SUPERCONDUTTORE CON TEMPERATURA DI TRANSIZIONE SUPERIORE A 77 KELVIN
$Bi_2Sr_2CuO_6$	$CaCu_2O_8$ o $Ca_2Cu_3O_{10}$	0-110	
$Tl_2Ba_2CuO_6$	$CaCu_2O_8$ o $Ca_2Cu_3O_{10}$	80-125	FRA I SUPERCONDUTTORI NOTI $Tl_2Ba_2Ca_2Cu_3O_{10}$ HA LA TEMPERATURA DI TRANSIZIONE PIÙ ALTA
$TlBa_2Cu_2O_5$	$CaCu_2O_7$, $Ca_2Cu_3O_9$ o $Ca_3Cu_4O_{11}$	0-122	
$Pb_2Sr_{2+x}Pr_{1-x}Cu_3O_8$	Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er o Tm	70-85	
$Bi_2Sr_2Sm_{2-2x}Ce_{2x}Cu_2O_{10}$	Eu o Gd	20-25	
$Ba_{1,33}Nd_{0,67}Sm_{1,33}Ce_{0,67}Cu_3O_9$	Sm, Eu o Gd Nd, Eu o Gd	40	
$La_{2-x}Sr_xCaCuO_6$	NESSUNA CONOSCIUTA	60	

0,17 questi materiali raggiungono la loro massima temperatura di transizione, circa 25 kelvin. Questi composti continuano a essere oggetto di molte ricerche, poiché possiedono singolari caratteristiche chimiche, non ancora comprese appieno, che possono influire sulla natura dei portatori di carica.

I superconduttori con le più alte temperature di transizione conosciute furono scoperti nel 1988 da Allen M. Hermann e colleghi dell'Università dell'Arkansas e sono composti di tallio, bario, calcio, rame e ossigeno. La tossicità dell'ossido di tallio impone la necessità di maneggiare questi materiali in condizioni di laboratorio molto ben controllate: ciò rende ancora dubbia la possibilità di un loro impiego commerciale. Nondimeno la struttura cristallina di questi composti ha confermato che per conseguire temperature di transizione elevate sono essenziali i piani definiti da atomi di rame e ossigeno.

In questi materiali il tallio si coordina con l'ossigeno formando un ottaedro di grandi dimensioni. Questi ottaedri definiscono un piano giacente sopra un piano formato da ottaedri o piramidi di rame e ossigeno. Una formula generale per i superconduttori contenenti tallio è $Tl_mBa_2Ca_{n-1}Cu_nO_{m+2n+2}$. Il deponente *m* indica il numero di strati di ottaedri di tallio e ossigeno; attualmente sono noti solo composti aventi uno o due strati. Il numero *n* di strati di atomi di rame può variare da uno a quattro. Tutti i composti della serie sono superconduttori ad alta temperatura e quello con la temperatura di transizione più alta, 125 kelvin, è $Tl_2Ba_2CaCu_3O_{10}$.

Benché i superconduttori a base di ossido di rame scoperti di recente abbiano una struttura più complessa dei loro predecessori, l'elemento essenziale per la superconduzione sono i piani di rame e ossigeno. Questi piani si alternano ad altri che fungono da distanziatori e, soprattutto, da riserve di cariche positive e negative. Lo stato elettronico di questi ultimi determina l'entità della carica sui piani di rame e ossigeno e la temperatura di transizione del composto.

Più di 15 anni fa, Arthur W. Sleight e collaboratori della Du Pont ottennero un materiale ceramico costituito da bario, piombo, bismuto e ossigeno. Benché avesse una temperatura di transizione di soli 12 kelvin, questo materiale fu il composto che spinse Bednorz e Müller sulle tracce dei superconduttori ceramici. Nel 1988 i miei colleghi e io agli AT&T Bell Laboratories scoprimmo che un composto a base di bario, bismuto e ossigeno era superconduttore alla sorprendente temperatura di 30 kelvin. Questi materiali hanno molto in comune con gli ossidi di rame e rappresentano di per sé un caso interessante. Forse essi, o altri materiali ancora ignoti, indicheranno nuove strade per la superconduttività ad alta temperatura: solo il tempo permetterà di saperlo.

BIBLIOGRAFIA

KINGERY W. D., BOWEN H. K. e UHLMANN D. R., *Introduction to Ceramics*, seconda edizione, John Wiley & Sons, Inc., 1976.

WEST ANTHONY R., *Solid State Chemistry and Its Applications*, John Wiley & Sons, Inc., 1984.

High-Temperature Superconductivity, ristampe da «Physical Review Letters» e «Physical Review B», The American Physical Society, 1987.

SLEIGHT A. W., *Chemistry of High-Temperature Superconductors* in «Science», 242, n. 4885, 16 dicembre 1988.

CAVA R. J., *Structural Chemistry and the Local Charge Picture of Copper Oxide Superconductors* in «Science», 247, n. 4943, 9 febbraio 1990.

VANDERAH TERRELL (a cura), *Chemistry of Superconductor Materials*, Noyes Publications, in stampa.

Infezioni opportunistiche correlate all'AIDS

In attesa di sconfiggere il virus dell'AIDS, le speranze di aumentare la sopravvivenza e migliorare la qualità della vita dei pazienti si baseranno soprattutto sullo sviluppo di terapie per le malattie associate alla sindrome

di John Mills e Henry Masur

Il virus dell'immunodeficienza dell'uomo (HIV) non causa di per sé la maggior parte delle patologie e dei decessi associati alla sindrome da immunodeficienza acquisita (AIDS). Questo agente virale può danneggiare direttamente alcuni organi, ma soprattutto, abbattendo progressivamente le difese dell'organismo, prepara la via allo sviluppo di infezioni opportunistiche: invasioni di microrganismi che proliferano in modo incontrollato a causa della deficiente funzionalità del sistema immunitario. Le infezioni di questo tipo, di rado patologiche nelle persone con sistema immunitario integro, sono responsabili anche del 90 per cento dei decessi nei pazienti affetti da AIDS (lo stadio finale dell'infezione da HIV).

I farmaci che si dimostrano attivi contro l'HIV, come la zidovudina (AZT), in realtà non eliminano completamente il virus, ma rallentano il declino del sistema immunitario da esso provocato, ritardando in questo modo la comparsa delle infezioni opportunistiche. Fino a quando non si troverà una vera e propria cura per l'infezione da HIV, tuttavia, la sopravvivenza e la qualità della vita dei pazienti dipenderanno soprattutto dalla messa a punto di terapie capaci di prevenire o curare più efficacemente le singole infezioni opportunistiche.

Le ricerche su queste terapie si sono ampliate significativamente nell'ultimo decennio e stanno già consentendo progressi nel trattamento dei pazienti. Per esempio, quando si manifestò l'epidemia di AIDS, non vi era alcun mezzo per combattere le infezioni da citomegalovirus, un virus erpetico che, nei soggetti immunodepressi, può causare cecità e danni all'apparato gastrointestinale o ai polmoni. Oggi, per la terapia in fase acuta e per la prevenzione delle recidive, disponiamo di un farmaco già autorizzato dalla Food and Drug Administration e di un secondo farmaco in avanzata fase di sperimentazione clinica. Fino a poco tempo fa la polmonite da *Pneumocystis carinii* colpiva almeno una volta l'85 per cento di tutti i pazienti affetti da AIDS. Attualmente più dei due terzi di questi episodi possono essere prevenuti con una terapia farmacologica.

Le infezioni opportunistiche correlate all'AIDS sono molto varie, tuttavia hanno in comune diverse caratteristiche. Per esempio, la maggior parte di esse colpisce anche persone immunodepresse per altre cause, come coloro che assumono farmaci immunosoppressori per prevenire il rigetto di organi trapiantati. (Ci sono, tuttavia, alcune eccezioni: per ragioni non ancora chiarite, certe patologie frequenti nei malati di AIDS sono rare in altri individui immunodepressi.)

Inoltre, le malattie infettive associate all'infezione da HIV sono provocate da organismi assai comuni, anche se la loro incidenza cambia da regione a regione al variare dei microrganismi endemici. Tali malattie di solito sono dovute alla riattivazione di infezioni quiescenti controllate dal sistema immunitario del paziente prima dell'infezione da HIV.

Di regola le infezioni che si instaurano nei pazienti sieropositivi nei confronti dell'HIV sono quelle di solito controllate dalla risposta mediata da cellule del sistema immunitario, risposta dovuta ai linfociti *T* e ai macrofagi, che vengono decimati da questo tipo di virus. Al contrario, le infezioni che sono contrastate principalmente dalla risposta immunitaria mediata da anticorpi o da difese aspecifiche (fra le quali ve ne sono molte provocate da batteri) sono di solito controllate abbastanza efficacemente dal sistema immunitario, almeno per un certo periodo di tempo.

Si conosce solo parzialmente il modo in cui l'HIV inibisce la risposta immunitaria mediata da cellule. Si sa che il virus elimina gradualmente l'insieme dei linfociti *T* che portano il recettore CD4 (linfociti *T* CD4+) e in effetti ci si dovrebbe attendere che una simile perdita indebolisca la risposta immunitaria mediata da cellule, poiché questi linfociti, noti anche come linfociti *T* coadiuvanti o *helper*, controllano l'attività di molte altre componenti della risposta immunitaria cellulare. Tuttavia, una disfunzione immunitaria è spesso presente anche prima che divenga evidente un declino dei linfociti *T* CD4+, il che significa che la perdita di queste cellule non è la sola causa del danno immunitario.

I ricercatori hanno scoperto inoltre che lo sviluppo di particolari infezioni opportunistiche è in relazione al numero di linfociti *T* CD4+ nel sangue. Gli individui sani hanno circa 1000 di queste cellule in un millimetro cubo di sangue, mentre negli individui sieropositivi la concentrazione diminuisce in media di 40-80 ogni anno.

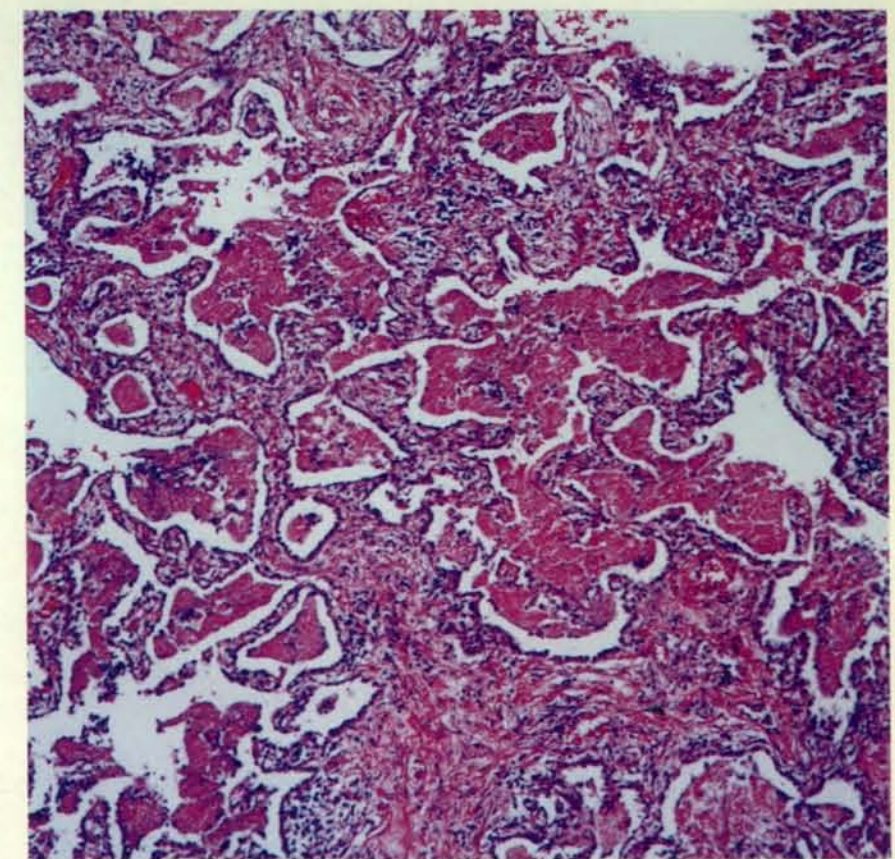
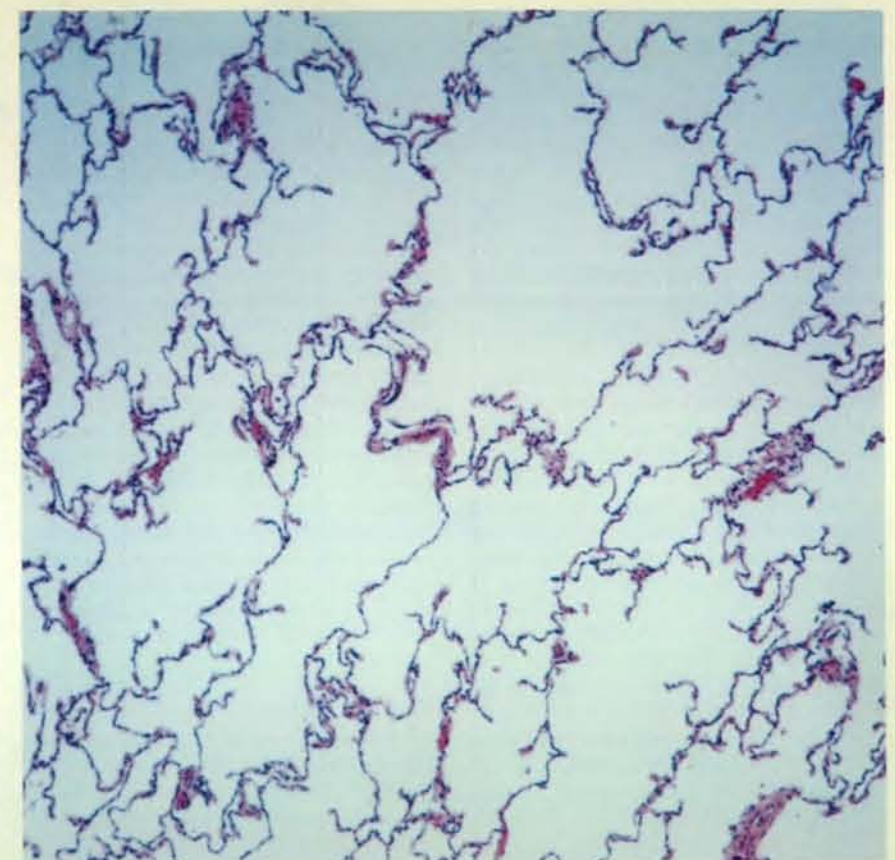
Quando il numero di questi linfociti scende tra 400 e 200 per millimetro cubo, compaiono di solito le prime infezioni, relativamente benigne ma fastidiose, che colpiscono la cute e le mucose. Tra queste possono esserci la candidiasi o mughetto (dolorose ulcerazioni biancastre nel cavo orale provocate da *Candida albicans*), infezioni cutanee e nervose prodotte dal virus varicella-zoster, il piede d'atleta in forma insolitamente grave e la leucoplachia orale vegetante (placche biancastre sulla lingua provocate dal virus di Epstein-Barr).

Allorquando compaiono questi sintomi, si è soliti dire che il soggetto è affetto dal complesso correlato all'AIDS, o ARC. Lo stesso vale per i pazienti che soffrono di febbri croniche criptogenetiche, diarrea, sudori notturni profusi o perdita di peso. Con il progressivo aggravarsi della deficienza immunitaria, si sviluppano di solito le infezioni opportunistiche gravi caratteristiche dell'AIDS. Queste comprendono spesso tre delle più comuni cause di morte per i malati di AIDS: la polmonite da *P. carinii*, la meningite criptococcica (provocata da un fungo) e la toxoplasmosi (una malattia protozoaria che interessa il sistema nervoso centrale). In passato, queste tre sole infezioni erano responsabili negli Stati Uniti del 50-70 per cento dei decessi di pazienti affetti da AIDS.

John P. Phair della Northwestern University e collaboratori hanno dimostrato che il rischio che si sviluppi una polmonite da *P. carinii* aumenta drasticamente dopo che il numero di linfociti *T* CD4+ è sceso al di sotto dei 200 per millimetro cubo: senza una terapia preventiva, più di un quinto dei pazienti con un numero di linfociti *T* CD4+ così ridotto incorrerà in un primo episodio di polmonite da *P. carinii* entro un anno. Il rischio di patologie provocate da altre particolari infezioni, come quella da citomegalovirus, non diventa significativo fino a quando il numero di linfociti *T* CD4+ non sia sceso al di sotto dei 100 per millimetro cubo.

L'approccio ideale al trattamento delle infezioni opportunistiche correlate all'AIDS presenta tre aspetti: la prevenzione dell'infezione (profilassi primaria), il trattamento dell'infezione acuta e infine la prevenzione delle recidive (profilassi secondaria). Oggi si pone molta attenzione alla profilassi secondaria, perché in mancanza di essa le infezioni opportunistiche si ripresentano quasi sempre. Questo avviene perché di rado le terapie antibiotiche riescono a eliminare completamente i microrganismi responsabili. Negli individui sani queste terapie sono efficaci in quanto danno all'organismo il tempo di organizzare le proprie difese contro l'agente patogeno, mentre nei pazienti con infezione da HIV il sistema immunitario compromesso non è in grado di eliminare l'agente patogeno.

Di solito ci si attende che un farmaco efficace contro un'infezione acuta possa essere utile anche nella profilassi primaria e in quella secondaria. Sfortunatamente, molti farmaci che possono essere tollerati per un breve periodo di tempo manifestano inconvenienti che ne sconsigliano la somministrazione prolungata a scopo preventivo, come una potenziale tossicità o la necessità di somministrazione endovenosa che, oltre a essere poco pratica e costosa, può dare origine a ulteriori infezioni. Questi problemi evidenziano l'importanza di mettere a punto una serie di farmaci per ciascuna delle principali infezioni opportunistiche caratteristiche dell'AIDS. È necessario avere a disposizione diverse alternative così che un paziente che non reagisca a un composto o non ne tolleri gli effetti collaterali o l'interazione con altri farmaci (come la zidovudina) possa far ricorso ad altre terapie.

Il trattamento in tre fasi delle principali infezioni opportunistiche correlate all'AIDS ha avuto successo soprattutto nel caso della polmonite da *P. carinii*, sebbene la strategia di sopravvivenza del microrganismo sia ancora sconosciuta. Questo agente patogeno, che usualmente infetta i polmoni (ma può colpire altri

In un individuo sano (*in alto*) gli alveoli polmonari contengono solamente gas, mentre appaiono riempiti di liquido e detriti cellulari (*in rosa*), che impediscono gli scambi gassosi e il trasporto di ossigeno verso gli organi vitali, nei pazienti affetti da polmonite da *Pneumocystis carinii* (*in basso*), una delle principali infezioni opportunistiche correlate all'AIDS.

organi), è spesso classificato come un protozoo, ma nuovi studi sembrano indicare che possa trattarsi di un fungo.

Probabilmente *P. carinii* si trasmette da un individuo all'altro tramite minuscole goccioline di saliva esalate nell'aria, allo stesso modo della tubercolosi o dell'influenza. Sebbene si pensi che nei polmoni della maggior parte degli adulti sia normalmente presente *P. carinii*, non è mai stato documentato un solo caso di polmonite attribuibile a questo microrganismo in individui con sistema immunitario integro.

Negli immunodepressi i microrganismi invadono gli alveoli, le unità funzionali del polmone dove avviene lo scambio di ossigeno e anidride carbonica. Nonostante il precario stato del sistema immunitario, l'organismo dei pazienti affetti da AIDS organizza una risposta infiammatoria all'infezione. Leucociti, anticorpi e altre proteine del sangue si accumulano negli alveoli e il tessuto polmonare appare infiltrato da globuli bianchi. Inoltre sia il tessuto, sia gli alveoli si riempiono di liquido impedendo gli scambi gassosi e quindi il trasporto di ossigeno verso gli organi vitali. Una polmonite da *P. carinii* non curata è quasi sempre fatale, ma anche con una terapia adeguata l'esito è infausto nel 10-20 per cento dei casi.

I medici usano trattare questa malattia con due farmaci. Uno è l'isetionato di pentamidina somministrato per via endovenosa. (La forma aerosol del prodotto, inalabile e molto pubblicizzata, ha una notevole utilità preventiva ed è promettente per il trattamento non tossico dell'infezione acuta, ma non ne è stata ancora definitivamente dimostrata l'efficacia per la terapia intensiva.) L'altro farmaco è il cotrimossazolo somministrato per via orale o endovenosa, che è una associazione di due altre sostanze, il sulfametossazolo e il trimetoprim.

Si ritiene che in certi microrganismi, e presumibilmente anche in *P. carinii*, la pentamidina prevenga la replicazione microbica inibendo la sintesi di DNA, anche se il suo meccanismo d'azione non è ancora stato chiarito. Anche il cotrimossazolo impedisce la sintesi del DNA (*si veda l'illustrazione a pagina 37*); il sulfametossazolo in esso contenuto inibisce infatti l'attività di un enzima - la diidropteroatosintetasi (DHPS) - che favorisce la produzione di un derivato dell'acido folico necessario per la produzione dei nucleotidi, gli elementi costitutivi del DNA. Gli inibitori della DHPS sono interessanti perché impediscono la sintesi del DNA microbico, ma non interferiscono in quella del DNA umano. Le cellule dell'uomo e degli altri mammiferi non necessitano di questo enzima (e quindi non lo possiedono) perché possono utilizzare i folati degli alimenti.

Il trimetoprim contenuto nel cotrimossazolo completa l'azione del sulfametossazolo inibendo l'attività di un enzima correlato alla DHPS, la diidrofolatoreduttasi (DHFR), che trasforma il folato sintetizzato dai microrganismi rendendolo direttamente utilizzabile nella produzione dei nucleotidi. Il trimetoprim inibisce anche la DHFR umana, ma solo in minima parte.

La pentamidina e il cotrimossazolo sono entrambi molto efficaci contro *P. carinii*, particolarmente se la terapia viene iniziata prima che si manifesti un sensibile aggravamento delle condizioni del paziente. Esistono tuttavia degli aspetti negativi. Entrambi i farmaci sono di scarsa utilità nei pazienti in condizioni molto gravi e producono sovente imponenti effetti collaterali. Circa la metà dei pazienti non è in grado di tollerare il primo farmaco somministrato e deve ricorrere a un preparato alternativo. La pentamidina introdotta per via endovenosa può compromettere la funzione renale o indurre una grave ipoglicemia. Il cotrimossazolo può causare grave nausea, eruzioni cutanee, febbre, diminuzione del numero di leucociti e piastrine ed epatopatia.

Molti ricercatori stanno mettendo a punto nuovi farmaci meno tossici e più efficaci di quelli esistenti e con modalità di assunzione meno rischiose, per esempio, della pentamidina. Un numero sempre maggiore di inibitori della DHFR e della DHPS, sia nuovi sia già noti, è in fase di sperimentazione clinica. Tra essi ci sono il trimetrexato (somministrato per via endovenosa) e il piritrexim (per via orale), due nuovi inibitori della DHFR.

Altri protocolli terapeutici che vengono sperimentati clinicamente con risultati promettenti contemplano l'associazione di due composti standard - la clindamicina (un antibiotico che inibisce la sintesi proteica batterica) e la primachina (un antimalarico) - oppure l'associazione di trimetoprim e dapsone (un solfone inibitore della DHPS). In corso di sperimentazione clinica sono anche la difluorometilornitina, o DFMO (un inibitore della sintesi del DNA in alcuni parassiti), e un agente chiamato provvisoriamente composto 566C80, che blocca l'attività dei mitocondri. I ricercatori stanno peraltro controllando *in vitro* l'efficacia contro *P. carinii* di migliaia di inibitori della DHPS e della DHFR.

Per contrastare in maniera più efficace l'infezione acuta vengono prese in considerazione anche terapie che non aggrediscono direttamente il microrganismo. Diversi studi clinici recenti indicano che la pronta somministrazione di glicocorticoidi riduce l'infiammazione polmonare e aumenta la possibilità di sopravvivenza dei pazienti in gravi condizioni. Dal momento che gli individui colpiti da *P. carinii* sono privi del surfattante naturale polmonare, una sostanza tensioattiva che mantiene aperti gli alveoli, in futuro potrebbe essere sperimentata una terapia sostitutiva con un tensioattivo artificiale somministrato per aerosol.

Per quanto riguarda la profilassi primaria e secondaria, la pentamidina inalata, il dapsone e il cotrimossazolo sembrano essere tutti molto efficaci. L'efficacia e la tossicità relative di questi composti, tuttavia, sono ancora in via di accertamento. Sulla base della scoperta di Phair, parrebbe che la profilassi primaria specifica per *P. carinii* debba iniziare non appena il numero di linfociti *T* CD4+ scende al di sotto dei 200 per millimetro cubo, o ancor prima se il paziente soffre di ARC o AIDS.

L'importanza della profilassi secondaria emerge chiaramente esaminando dati recenti. Fino a non molto tempo fa circa la metà di tutti i pazienti sopravvissuti a un primo episodio di polmonite da *P. carinii* subiva una recidiva nel giro di un anno, che assumesse o meno zidovudina. Oggi soltanto il 10-20 per cento dei soggetti trattati con zidovudina e un agente profilattico specifico per *P. carinii* mostra una ricomparsa dell'infezione nel giro di un anno. Inoltre, coloro che non ricevono una terapia preventiva specifica hanno una aspettativa di vita di un solo anno dopo il primo episodio di polmonite; affiancando una profilassi secondaria alla zidovudina la sopravvivenza media sale a circa tre anni.

Questi dati sono incoraggianti; siamo tuttavia in molti a sperare che la disponibilità di nuovi farmaci per il trattamento dell'infezione acuta da *P. carinii* consenta di mettere a punto terapie preventive efficaci, non tossiche, scevre di effetti collaterali e poco costose.

I progressi nello sviluppo di nuove terapie per altre infezioni opportunistiche correlate all'AIDS sono stati più lenti. Un caso eloquente è quello della toxoplasmosi. La malattia è provocata da *Toxoplasma gondii*, un parassita unicellulare i cui ospiti sono alcuni piccoli mammiferi e uccelli e l'uomo. Può venire trasmesso all'uomo tramite il contatto con feci di animali infetti o attraverso il consumo di carni infette poco cotte. Negli Stati Uniti circa il 20 per cento della popolazione adulta ospita il parassita, di solito nel tessuto cerebrale o muscolare.

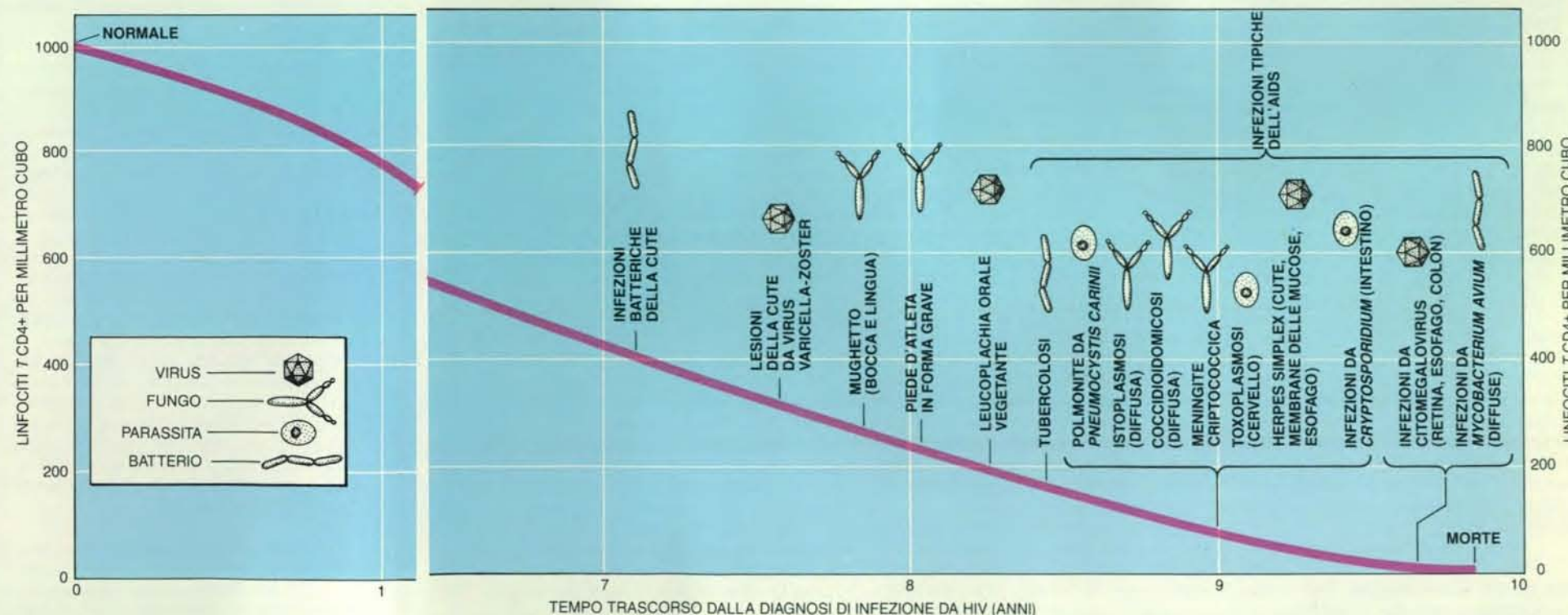
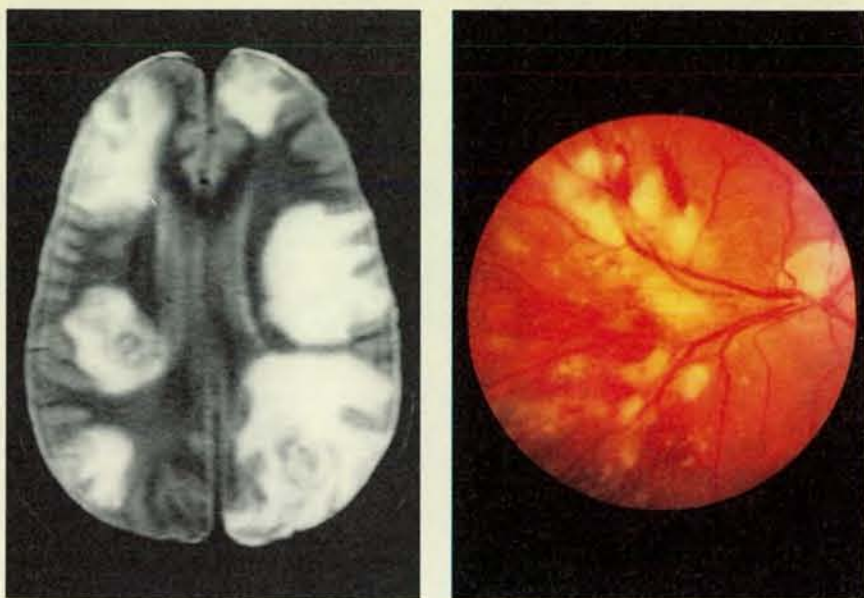
L'infezione latente si attiva, in genere nell'encefalo, nel 5-15 per cento dei pazienti affetti da AIDS. I parassiti si riproducono velocemente e inducono una risposta infiammatoria che distrugge le cellule cerebrali, provocando sintomi quali convulsioni, confusione mentale ed emiparesi.

Per ora l'unica terapia efficace per la toxoplasmosi è un'associazione di pirimetamina (un inibitore della DHFR) e sulfadiazina (un inibitore della DHPS) somministrabile oralmente. Gran parte dei pazienti mostra un miglioramento nel giro di una-tre settimane, via via che lo sviluppo dei microrganismi viene impedito e l'infiammazione del tessuto cerebrale diminuisce: purtroppo i danni alle cellule cerebrali sono per lo più irreversibili. Come nel caso della polmonite da *P. carinii*, è indispensabile la profilassi secondaria, poiché i farmaci prescritti per la toxoplasmosi reprimono l'infezione, ma non la guariscono. La terapia provoca molti degli effetti collaterali già segnalati per il cotrimossazolo e infatti circa il 30 per cento dei pazienti non tollera un ciclo di trattamento completo.

Fortunatamente, vi sono alcune alternative promettenti. In una sperimentazione clinica a scala ridotta (e non ancora ripetuta) un'associazione di pirimetamina, come inibitore della DHFR, e di clindamicina, come agente antibatterico, si è dimostrata quasi altrettanto efficace dell'associazione standard pirimetamina-sulfadiazina. Oltre a ciò, la ricerca di altri antibiotici dotati di attività contro *T. gondii*, *in vitro* o nei topi, ha permesso di identificare un certo numero di possibili candidati. Tra di essi vi sono analoghi della clindamicina e della tetraciclina, oltre ad alcuni inibitori della DHFR e della DHPS; nessuno di questi composti, tuttavia, è mai stato studiato a fondo nell'uomo.

Seguendo una diversa strategia, alcuni ricercatori cercano di sfruttare il fatto che *T. gondii* può sopravvivere solo all'interno delle cellule, tra le quali preferisce i macrofagi. Negli individui con sistema immunitario normale i macrofagi, dopo aver ingerito i microrganismi invasori, vengono attivati per distruggerli. I fattori di attivazione sono l'interferone gamma e altre citochine secrete dai linfociti *T*. Ma, a causa della distruzione dei linfociti *T*, l'interferone gamma viene prodotto in piccola quantità nei soggetti con infezione da HIV, sicché i macrofagi ingeriscono il parassita, ma non lo distruggono. Alcuni ricercatori stanno valutando i possibili vantaggi dell'aggiunta di interferone gamma al trattamento standard con pirimetamina e sulfadiazina. In colture di cellule e in animali da laboratorio questo metodo ha migliorato l'efficacia della terapia combinata e si prevede che la sperimentazione clinica inizi fra breve tempo.

Il trattamento con pirimetamina e sulfadiazina viene di norma proseguito per tutta la vita del paziente onde prevenire recidive. Si discute, tuttavia, sull'opportunità di effettuare in tutti i pazienti la profilassi primaria. Negli Stati Uniti, molti medici esitano a somministrare farmaci potenzialmente tossici a tutti i sieropositivi per l'HIV in quanto, senza trattamento, solo un quinto o meno di essi è destinato ad ammalarsi di toxoplasmosi. In Francia, dove questa malattia è molto più comune, la profilassi primaria è invece raccomandata con frequenza. Studi per valutare i possibili vantaggi di queste terapie sono in corso sia negli Stati Uniti sia in Europa.

Al contrario della toxoplasmosi, le più comuni infezioni batteriche che affliggono gli adulti affetti da AIDS e che

Le infezioni opportunistiche che colpiscono con frequenza i soggetti contagiati da HIV, il virus responsabile dell'AIDS, si presentano in genere in una successione correlata alla concentrazione ematica dei linfociti *T* CD4+ o linfociti *helper*. Il numero di queste cellule, che diminuisce nel tempo, rispecchia il grado di compromissione del sistema immunitario causata dal virus. Il grafico è stato costruito a partire da dati raccolti in gran parte da Suzanne M. Crowe del Fairfield Hospital di Melbourne, in Australia.

Alcune infezioni sono diagnosticabili sulla base di immagini cliniche. Per esempio, la toxoplasmosi, che può provocare convulsioni, produce nel cervello lesioni delimitate che appaiono come aree chiare nelle immagini ottenute con il metodo della risonanza magnetica (*a sinistra*). In questa fotografia di retina danneggiata da citomegalovirus (*a destra*) le zone giallastre rappresentano il tessuto necrotico, mentre le chiazze rosse (che coprono i vasi sanguigni) indicano un'emorragia nella retina. (La zona scura, che corrisponde a un distacco della retina dal fondo dell'occhio, è in parte responsabile dell'emorragia.)

sono causate da organismi del ceppo *Mycobacterium avium* raramente colpiscono il cervello; esse però danneggiano molti altri organi e tessuti, in particolare i polmoni, il fegato, la milza, l'intestino, i linfonodi, il midollo osseo e il sangue. In effetti di solito sono disseminate in tutto l'organismo.

Questi batteri, che sono strettamente correlati all'agente responsabile della tubercolosi (*M. tuberculosis*), sono diffusissimi nell'ambiente, per esempio nella polvere, nel terreno, negli animali da cortile e nei prodotti caseari. Presumibilmente essi penetrano nell'organismo con l'inalazione di polvere e di minuscole goccioline o tramite il cibo e l'acqua. Nonostante la loro enorme diffusione, raramente diventano virulenti, anche in persone immunodepresse per cause diverse dall'AIDS.

Tuttavia quasi il 50 per cento dei pazienti affetti da AIDS presenta infezioni diffuse al momento del decesso. Negli individui immunologicamente integri, i micobatteri inducono un'energica risposta immunitaria mediata da cellule; nei malati di AIDS i tessuti contengono quantità enormi di *M. avium*, ma l'infiammazione appare minima.

In effetti, i micobatteri proliferanti potrebbero essere la causa dei sintomi che spesso accompagnano l'infezione, quali febbre, sudorazione, perdita di peso e affaticamento, ma non lo si è mai determinato con certezza: nei pazienti con AIDS sono infatti generalmente in corso anche infezioni causate da molti altri microrganismi (come il citomegalovirus e lo stesso HIV). Così pure, la diminuzione di globuli rossi, globuli bianchi e piastrine che può accompagnare l'infezione da *M. avium* potrebbe essere causata dall'invasione del midollo osseo da parte dei micobatteri. D'altro canto, in alcuni pazienti che presentano enormi quantità di micobatteri in tutto l'organismo non si riscontrano quasi segni di disfunzioni specifiche.

Purtroppo, il farmaco di elezione contro la tubercolosi, l'isoniazide, invariabilmente non riesce a controllare la proliferazione di *M. avium*. I ricercatori hanno perciò sperimentato un gran numero di altri prodotti, alcuni dei quali, da soli o in associazione, sono in grado di limitare la crescita di questi batteri in colture cellulari o in topi immunodepressi. Fra i composti spesso prescritti per la tubercolosi, quelli classificati come amminoglicosidi (quali la streptomicina e la amikacina), oltre ai farmaci del gruppo delle rifamicine e all'etambutolo, hanno tutti mostrato una certa attività contro *M. avium*. La stessa cosa vale per alcuni antibiotici: la ciproflossacina, l'imipenem e la clofazimina.

Due sperimentazioni cliniche distinte hanno dimostrato che associazioni di quattro o cinque di questi farmaci possono ridurre la quantità di batteri in circolo e alleviare sintomi quali la febbre. Se ulteriori studi confermeranno queste scoperte, associazioni di queste sostanze potrebbero essere utilizzate anche nella profilassi primaria e secondaria.

Anche le ricerche sui trattamenti per le principali infezioni fungine e virali correlate all'AIDS stanno compiendo progressi. Per esempio, è stato sviluppato un nuovo protocollo terapeutico per la meningite criptococcica - che colpisce circa il 10 per cento dei pazienti con AIDS negli Stati Uniti - mentre altri sono in fase sperimentale.

L'organismo responsabile della meningite criptococcica, il lievito *Cryptococcus neoformans*, assomiglia strettamente al fungo che causa il mughetto, ma con una importante differenza: *C. neoformans* si diffonde molto più facilmente attraverso la cute e le mucose. Il suo rivestimento mucillaginoso fa sì che macrofagi, granulociti e altre cellule del sistema immunitario lo fagocitino con difficoltà e che, se ingerito, possa resistere ai danni provocati dagli enzimi di quelle cellule. Ciò che rimane un mistero è il meccanismo patogenetico di *C. neoformans*; finora non sono state identificate sostanze tossiche prodotte da questo organismo.

Si ritiene che l'infezione inizi nei polmoni, dove nei soggetti in condizioni normali di salute resterebbe probabilmente confinata. Nei pazienti immunodepressi, tuttavia, il microrganismo invade il circolo sanguigno per poi insediarsi nelle meningi, le membrane che rivestono l'encefalo. L'invasione, potenzialmente letale, delle meningi è denunciata da forti dolori al capo, rigidità della nuca e spesso febbre.

Per 30 anni, e fino a tempi recenti, l'anfotericina B è stato l'unico trattamento disponibile. Questo farmaco si lega all'ergosterolo, una sostanza simile al colesterolo che si trova solo nella membrana cellulare dei funghi. Il legame danneggia la membrana e la fuoriuscita del materiale intracellulare provoca la morte della cellula.

Purtroppo il farmaco, che deve essere somministrato per via endovenosa, danneggia i globuli rossi dell'uomo e può avere effetti nocivi sui reni. Oltre a ciò, la formulazione necessaria per rendere il composto attivo adatto alla somministrazione endovenosa produce in molti casi sintomi simili a quelli di una forte influenza, come brividi e febbre alta. Anche dopo un ciclo completo di terapia, nel 20 per cento circa dei malati di AIDS affetti da meningite criptococcica l'esito è infausto.

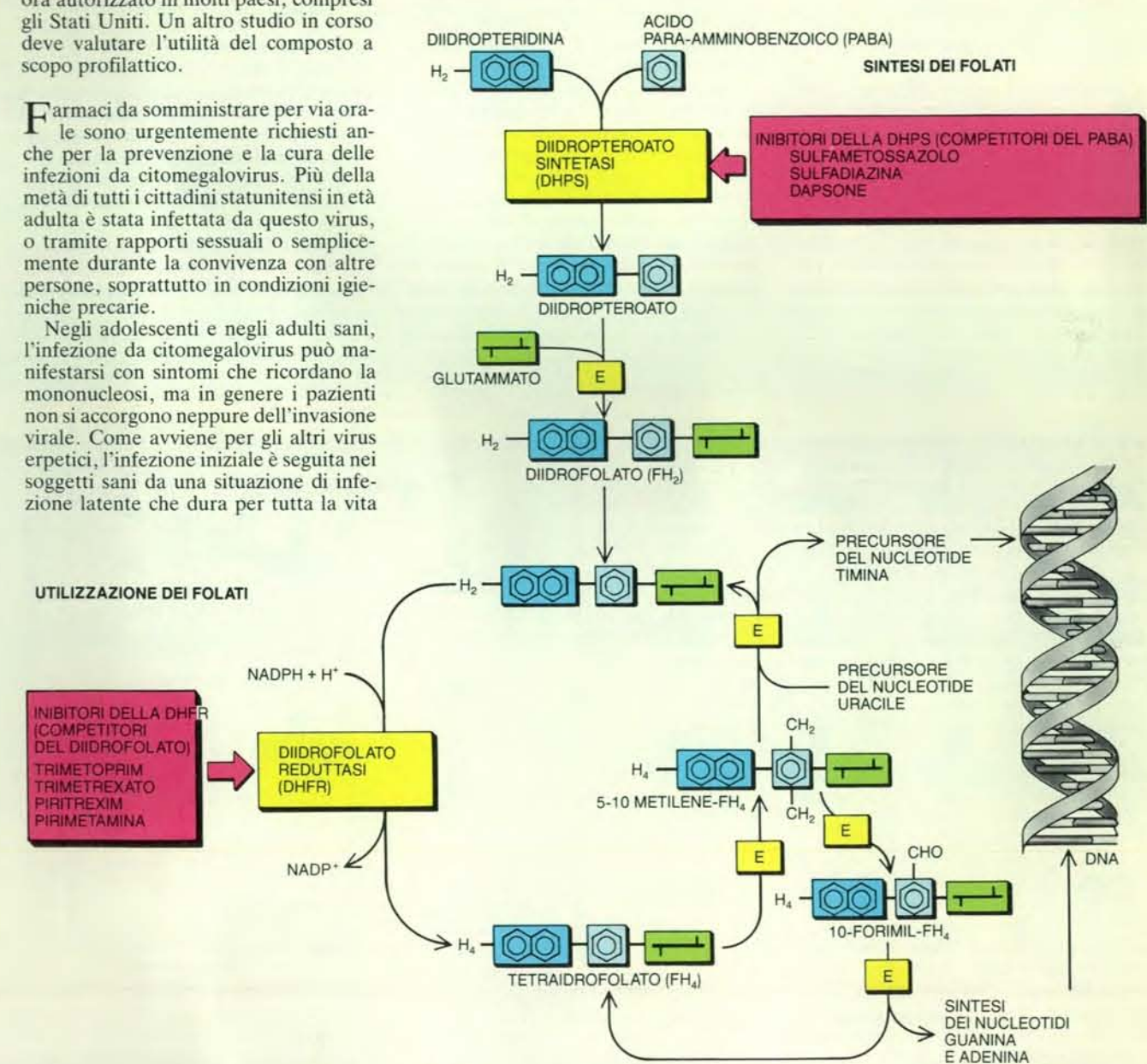
In teoria, l'anfotericina B potrebbe realmente eliminare le infezioni da funghi e non solo rallentarne il decorso, ma gli alti dosaggi richiesti finirebbero probabilmente per uccidere anche il paziente. Dato perciò che con i dosaggi somministrati l'infezione non viene eliminata, i pazienti che sopravvivono devono, anche in questo caso, ricevere una terapia preventiva per il resto della vita, sottoponendosi a iniezioni endovenose una o due volte alla settimana.

Un farmaco alternativo che possa essere assunto oralmente rappresenterebbe un notevole progresso per la profilassi. I composti che nelle prove cliniche si sono dimostrati più promettenti appartengono a un gruppo di antifungini, gli imidazoli, che inibiscono la sintesi dell'ergosterolo; due di essi - l'itraconazolo e il fluconazolo - suscitano particolare speranza. Le prime prove cliniche con l'itraconazolo sembrano indicare un'efficacia quasi pari a quella dell'anfotericina B, ma una minore tossicità. L'analisi di uno studio ad ampio raggio recentemente completato fa pensare che il fluconazolo possa essere efficace quanto l'anfotericina B, tanto che il farmaco è ora autorizzato in molti paesi, compresi gli Stati Uniti. Un altro studio in corso deve valutare l'utilità del composto a scopo profilattico.

Farmaci da somministrare per via orale sono urgentemente richiesti anche per la prevenzione e la cura delle infezioni da citomegalovirus. Più della metà di tutti i cittadini statunitensi in età adulta è stata infettata da questo virus, o tramite rapporti sessuali o semplicemente durante la convivenza con altre persone, soprattutto in condizioni igieniche precarie.

Negli adolescenti e negli adulti sani, l'infezione da citomegalovirus può manifestarsi con sintomi che ricordano la mononucleosi, ma in genere i pazienti non si accorgono neppure dell'invasione virale. Come avviene per gli altri virus erpetici, l'infezione iniziale è seguita nei soggetti sani da una situazione di infezione latente che dura per tutta la vita senza dare manifestazioni patologiche.

Nei malati di AIDS l'agente patogeno può essere sovente rilevato in grande quantità nel sangue e in molti altri tessuti. I danni funzionali, tuttavia, sono di solito limitati all'occhio e all'apparato gastrointestinale. Nel 7 per cento circa delle persone affette da AIDS il virus invade i coni e i bastoncelli, che sono le cellule fotorecettrici della retina. L'infezione finisce per uccidere le cellule colpite, danneggiando irreparabilmente la visione. Inoltre, è probabile che il distacco delle zone necrotiche della retina e la conseguente emorragia provochino cecità completa e improvvisa.

In un altro 5 per cento di malati di AIDS il virus si annida nell'apparato digerente, in un punto qualsiasi dalla bocca al retto, anche se colpisce più frequentemente l'esofago e il colon. Le ulcere nell'esofago rendono dolorosa la deglutizione e i dolori si possono irradiare anche al torace; nel colon danno spesso luogo a dolori addominali e a diarrea. Talvolta le emorragie incontrollabili che

Molti farmaci che si sono dimostrati efficaci contro la polmonite da *Pneumocystis carinii* o la toxoplasmosi (*nei riquadri rossi*) interferiscono con l'attività di due enzimi presenti nei microrganismi responsabili, la diidropteroatosintetasi (DHPS) e la diidrofolatoreduttasi (DHFR). Questi enzimi prendono parte alla sintesi(*in alto*) e all'utilizzazione (*in basso*) dei folati (derivati dell'acido folico) per la produzione di nucleotidi, i costituenti del DNA. Gli inibitori della DHPS assomigliano all'acido para-amminobenzoico (PABA), una subunità del prodotto principale della sintesi dei folati: il diidrofolato. Legandosi alla DHPS, i farmaci impediscono all'enzima di interagire con il PABA e quindi di legare quest'ultimo alla diidropteridina, un'altra subunità del diidrofolato. Gli inibitori della DHFR sono analoghi del diidrofolato. Competendo per il legame con la DHFR, essi impediscono la conversione del vero folato in una forma che possa essere a sua volta utilizzata nella sintesi dei nucleotidi. La lettera E nel riquadro rappresenta altri enzimi.

si verificano nei tessuti danneggiati dell'esofago e del colon possono condurre a morte il paziente.

I farmaci attualmente disponibili per il trattamento delle patologie da citomegalovirus sono tossici e devono essere somministrati per via endovenosa. Uno di essi, il ganciclovir, è approvato sia dalla FDA negli Stati Uniti sia in diversi altri paesi; un secondo, il foscarnet, è ancora in fase sperimentale, ma si prevede che verrà autorizzato fra breve tempo. Entrambi i composti inibiscono selettivamente la DNA-polimerasi virale, un enzima essenziale per la sintesi di nuovo DNA dai nucleotidi e quindi anche per la replicazione virale. Per altri aspetti, tuttavia, i due farmaci sono molto diversi.

Il ganciclovir, che è efficace contro molti virus erpetici, è un analogo di un nucleoside, la guanosina (un precursore di un nucleotide), e proprio la sua capacità di imitare la guanosina ne spiega l'attività antivirale. Le cellule infettate dal virus addizionano gruppi fosfato al farmaco come se fosse guanosina, formando ganciclovir trifosfato, un finto nucleotide. Questo compete nel formare legami con la DNA-polimerasi virale e la inattiva impedendole di legare i veri nucleotidi. Purtroppo il ganciclovir viene fosforilato in una certa misura anche nelle cellule non infettate e può quindi interferire con l'attività della DNA-polimerasi umana.

L'altro farmaco, il foscarnet, è un analogo di due fosfati che la DNA-polimerasi elimina dai nucleotidi trifosfati prima di addizionarli a una catena di DNA in formazione. Anch'esso viene assunto dalla DNA-polimerasi virale e ne impedisce il legame con i nucleotidi. Il foscarnet è attivo contro la DNA-polimerasi di tutti i virus erpetici che infettano l'uomo e anche contro la DNA-polimerasi dello stesso HIV; non è invece attivo contro l'enzima umano.

Interferendo con la replicazione virale, entrambi i farmaci arrestano il decorso della patologia retinica causata dal citomegalovirus e sembra che blocchino la diffusione dell'infezione anche in altri siti. Ma, dato che non eliminano realmente il virus, devono, come al solito, essere assunti con continuità. Per la profilassi secondaria il ganciclovir e il foscarnet devono essere somministrati una o più volte al giorno, il che richiede l'installazione di un catetere collegato a una vena del torace. Sebbene questa soluzione consenta la terapia, rende anche il paziente suscettibile a infezioni dei tessuti circostanti il catetere.

Un altro aspetto negativo è la tossicità del ganciclovir nei confronti dei globuli bianchi, in particolare dei granulociti. Nel 10-20 per cento dei pazienti la diminuzione dei granulociti è così grave da richiedere la sospensione del trattamento (dato che un basso livello di granulociti rende il paziente vulnerabile a infezioni batteriche talvolta mortali). Il foscarnet può causare disfunzioni renali e anomalie metaboliche e neurologiche, tutte fortunatamente reversibili.

Particolarmente preoccupante è la capacità dimostrata dal citomegalovirus di sviluppare resistenza al ganciclovir. Il problema non è ancora diffuso, ma diversi pazienti che erano stati trattati con questo farmaco a un certo punto non hanno mostrato più alcuna risposta. Per fortuna, nelle prove in laboratorio e in sperimentazioni cliniche preliminari, i ceppi resistenti al ganciclovir si sono rivelati sensibili al foscarnet. In laboratorio si è tuttavia riusciti a produrre ceppi mutanti di citomegalovirus che sono resistenti al foscarnet; nessuna di queste varianti è stata finora identificata nei pazienti, ma sembra sia solo una questione di tempo.

I farmaci contro i virus erpetici (*all'estrema sinistra*) interferiscono con la capacità delle DNA-polimerasi virali di costruire nuovo DNA dai nucleotidi. Questi enzimi normalmente rimuovono due dei tre gruppi fosfato di un nucleotide e poi collegano quanto resta a un gruppo OH del nucleotide aggiunto in precedenza. L'aciclovir, il ganciclovir e la vidarabina sono tutti fosforilati dagli enzimi nelle cellule infettate dal virus, e quindi assomigliano a normali nucleotidi con tre gruppi fosfato (*riquadro di sinistra*); il foscarnet simula il gruppo difosfato normalmente rimosso dai nucleotidi a opera di una polimerasi. L'aciclovir trifosfato può bloccare la produzione del DNA (*a destra*) facendo terminare la catena (*a*): se viene aggiunto a un filamento di DNA in formazione, ne arresta la crescita perché manca del gruppo OH necessario per l'addizione del nucleotide successivo. L'aciclovir trifosfato - come il ganciclovir e la vidarabina fosforilati - può anche legarsi fortemente alla polimerasi, bloccandone il sito di legame con il nucleotide e impedendo all'enzima di svolgere la sua attività con i veri nucleotidi (*b*). Anche il foscarnet agisce bloccando il sito di legame con il nucleotide (*c*).

A causa del fenomeno della resistenza e dei problemi posti dalla somministrazione per via endovenosa, vi è urgente necessità di terapie alternative. Abbiamo buone speranze che nel prossimo futuro vengano messi a punto uno o più analoghi del ganciclovir da somministrare oralmente. Uno di questi farmaci, lo HOE-602 (così chiamato dal nome della ditta produttrice, la società Hoechst) è in fase di sperimentazione clinica nella Germania Federale. La sostanza viene facilmente assorbita dall'apparato digerente e trasformata in ganciclovir nell'organismo.

Anche un altro analogo dei nucleosidi - l'aciclovir - potrebbe rivelarsi utile. Come il ganciclovir, esso inibisce la DNA-polimerasi virale, ma è notevolmente meno tossico e può essere assunto per via orale. D'altro canto, l'aciclovir è anche molto meno attivo del ganciclovir nei confronti del citomegalovirus. Ciononostante, alti dosaggi del farmaco si sono dimostrati in grado di sopprimere l'attivazione dell'infezione da citomegalovirus in pazienti sottoposti a trapianto e si ritiene che potrebbero avere lo stesso effetto anche nei malati di AIDS.

Intanto continua la ricerca di farmaci che siano non solo meno tossici del ganciclovir e del foscarnet, ma anche ugualmente o più efficaci nell'inibire la DNA-polimerasi virale. Nello stesso tempo si tenta di mettere a punto prodotti che interferiscano con funzioni virali diverse dalla replicazione del DNA, nella speranza che questi agenti siano attivi contro i virus resistenti al ganciclovir e al foscarnet.

Contrariamente a quanto avviene con il citomegalovirus, due altri virus erpetici rilevanti nell'AIDS - il virus dell'herpes simplex di tipo 1 e quello dell'herpes simplex di tipo 2 - provocano infezioni ricorrenti anche in molti individui con sistema immunitario integro. In genere questi virus provocano, rispettivamente, eruzioni circoscritte intorno alla bocca e ulcerazioni ricorrenti nella regione genitale. Nel contesto dell'infezione da HIV queste lesioni non sono circoscritte, ma invece di guarire tendono a ingrandirsi progressivamente e, dato che sono dolorose, possono pregiudicare l'appetito qualora si presentino intorno alla bocca.

Sebbene l'aciclovir non sia estremamente efficace contro il citomegalovirus, si è dimostrato una terapia eccellente contro i virus dell'herpes simplex. Anche nei pazienti con vaste ulcerazioni si ottiene un soddisfacente controllo delle lesioni nel giro di poche settimane. Purtroppo, sembra che anche i virus dell'herpes simplex stiano sviluppando una resistenza sempre maggiore al farmaco.

Gran parte dei ceppi resistenti riesce a sfuggire alla distruzione da parte dell'aciclovir in quanto non produce quantità significative di un enzima, la timidinachinasi. La versione virale dell'enzima provoca la fosforilazione dell'aciclovir, uno stadio fondamentale nell'attività di tutti i farmaci analoghi a nucleosidi. L'enzima virale non è tuttavia essenziale per la replicazione dei virus dell'herpes simplex, che sono in grado di compensare la carenza di questo enzima.

Una strategia logica per far fronte ai ceppi carenti di timidinachinasi consiste nel somministrare un analogo di un nucleoside che possa essere fosforilato dagli enzimi umani nelle cellule infettate. Uno di questi farmaci è la vidarabina, che da più di 10 anni viene utilizzata per curare infezioni gravi da virus dell'herpes simplex e varicella-zoster. Almeno *in vitro* molti ceppi di virus dell'herpes simplex resistenti all'aciclovir si sono dimostrati sensibili alla vidarabina. Questa scoperta fa ritenere che questo farmaco possa dimostrarsi utile dove l'aciclovir fallisce, anche se deve essere somministrato per via endovenosa.

Dato che il foscarnet non richiede fosforilazione per essere attivato, potrebbe anch'esso rivelarsi efficace e le prove di laboratorio e gli studi clinici preliminari sembrano incoraggianti. Tuttavia si sono già osservati ceppi di virus dell'herpes simplex resistenti al foscarnet.

Come dimostra il problema dei ceppi virali resistenti ai farmaci, le difficoltà da superare sono formidabili. Oltre a farmaci di facile somministrazione, non tossici ed economici da utilizzare a scopo profilattico devono essere messi a punto diversi trattamenti efficaci per ciascuna delle principali infezioni opportunistiche che colpiscono i malati di AIDS. In caso contrario, le infezioni e altre patologie correlate all'AIDS (come certe forme di tumore) che non possono essere messe sotto controllo prenderanno semplicemente il posto delle infezioni che vengono curate con efficacia.

Sono necessarie terapie anche per infezioni opportunistiche meno comuni (come l'infezione da criptosporidio, un parassita che può provocare diarrea) o diffuse solo in regioni limitate del globo. Alcuni esempi sono l'istoplasmosi e la coccidioidomicosi, entrambe infezioni fungine diffuse che si riscontrano, generalmente, la prima negli Stati Uniti centrali (soprattutto nella valle del Mississippi) e in poche altre regioni sparse in tutto il globo e la seconda nei deserti degli Stati Uniti sudoccidentali e in America Centrale e Meridionale.

La ricerca di nuove terapie è ben lungi dall'essere completata; tuttavia siamo soddisfatti di constatare che le patologie opportunistiche correlate all'infezione da HIV vengono sempre più considerate un obiettivo importante per lo sviluppo e lo studio di nuovi farmaci. Ogni progresso rappresenterà una significativa differenza nella qualità della vita di chi deve misurarsi ogni giorno con gli effetti diretti e indiretti del micidiale virus dell'AIDS.

BIBLIOGRAFIA

DEVITA VINCENT T., Jr., HELLMAN SAMUEL e ROSENBERG STEVEN A. (a cura), *AIDS: Etiology, Diagnosis, Treatment and Prevention*, seconda edizione, J. B. Lippincott Co., 1988.

LEOUNG GIFFORD e MILLS JOHN (a cura), *Opportunistic Infections in Patients with the Acquired Immunodeficiency Syndrome*, Marcel Dekker, 1989.

MURRAY JOHN F. e MILLS JOHN, *Pulmonary Infectious Complications of Human Immunodeficiency Virus Infection* in «American Review of Respiratory Disease», parte I, 141, n. 5, maggio 1990; parte II, 141, n. 6, giugno 1990.

Il linguaggio dei frattali

Queste strutture inconcepibilmente ricche di particolari sono molto di più che una semplice curiosità matematica: la geometria frattale è in grado di descrivere in modo compendioso oggetti e processi naturali complessi

di Hartmut Jürgens, Heinz-Otto Peitgen e Dietmar Saupe

«La natura ha fatto uno scherzo ai matematici. Può darsi che ai matematici dell'Ottocento difettasse l'immaginazione, ma alla natura no. Le stesse strutture patologiche che i matematici inventarono per affrancarsi dal naturalismo ottocentesco si rivelarono inerenti agli oggetti familiari da cui siamo circondati.»

FREEMAN DYSON
Characterizing Irregularity
in «Science», 12 maggio 1978

Le «strutture patologiche» ideate dai matematici dell'Ottocento hanno assunto negli ultimi anni la forma di frattali, figure matematiche dotate di dimensione frazionaria e non intera come accade per le ordinarie figure geometriche (per esempio le rette, che hanno dimensione uno, o i piani, che hanno dimensione due). Il fascino che esercitano oggi i frattali è dovuto in ampia misura al lavoro di Benoit B. Mandelbrot, del Thomas J. Watson Research Center della IBM a Yorktown Heights, nello Stato di New York. Mandelbrot coniò il termine frattale nel 1975, traendolo dal latino *fractus*, da *frangere*, cioè «rompere». Fu nel 1983 che il concetto di frattale acquisì vastissima notorietà presso i matematici, gli scienziati e il pubblico non specializzato, con la pubblicazione dell'opera pionieristica *The Fractal Geometry of Nature* dello stesso Mandelbrot.

I frattali sono molto di più che una semplice curiosità matematica: infatti essi offrono un metodo assai conciso per descrivere oggetti e formazioni. Molte strutture hanno una regolarità geometrica soggiacente, detta invarianza rispetto al cambiamento di scala o autosomiglianza. Se si esaminano questi oggetti a scale diverse si incontrano sempre gli stessi elementi fondamentali. Questa configurazione ripetitiva definisce la dimensione frazionaria, o frattale, della struttura. La geometria frattale sembra descrivere le forme e le configurazioni naturali in modo più succinto ed esteticamente più valido rispetto alla geometria euclidea tradizionale.

L'invarianza di scala trova un notevole parallelo nella teoria contemporanea del caos, nella quale molti fenomeni, benché seguano rigide regole deterministiche, si rivelano imprevedibili in linea di principio. Gli eventi caotici, come la turbolenza atmosferica o le pulsazioni cardiache, manifestano andamenti simili su scale temporali diverse, più o meno come gli oggetti dotati di autosomiglianza presentano forme strutturali simili su scale spaziali diverse. La corrispondenza tra frattali e caos non è accidentale; è viceversa il segno di una relazione profonda: la geometria frattale è la geometria del caos.

Un'altra analogia tra geometria frattale e teoria del caos consiste nel fatto che in entrambi i campi le scoperte più recenti sono avvenute grazie alla potenza dei calcolatori moderni. Questi progressi mettono in discussione la visione tradizionale della matematica. Molti matematici hanno salutato l'avvento dei calcolatori come un segno di rinnovamento e di liberazione, ma altri vedono in essi un allontanamento dalla matematica pura.

I frattali sono in primo luogo e soprattutto un linguaggio della geometria; tuttavia i loro elementi fondamentali non possono essere osservati direttamente. Sotto questo profilo essi presentano una differenza basilare rispetto ai ben noti elementi della geometria euclidea, come la retta e il cerchio. I frattali non si esprimono mediante forme primarie, bensì mediante algoritmi, vale a dire insiemi di procedure matematiche. Questi algoritmi vengono poi tradotti in forme geometriche con l'ausilio di un calcolatore. Dal momento che la riserva di elementi algoritmici è inesauribile, quando ci si sia impadroniti del linguaggio frattale si può descrivere la forma di una nube con la stessa precisione e semplicità con cui un architetto può descrivere una casa mediante una pianta tracciata nel linguaggio della geometria tradizionale.

Questa metafora è particolarmente appropriata per le idee fondamentali della geometria frattale. Le lingue indoeuropee sono basate su alfabeti finiti (per esempio le 26 lettere con cui si scrivono le parole inglesi o le 21 lettere della lingua italiana). Le lettere non hanno significato fino a quando non sono giustapposte a formare parole. Analogamente, la geometria euclidea è costituita solo da pochi elementi (la retta, il cerchio e così via), con i quali si possono costruire oggetti complessi che, in un certo senso, solo allora hanno un significato geometrico.

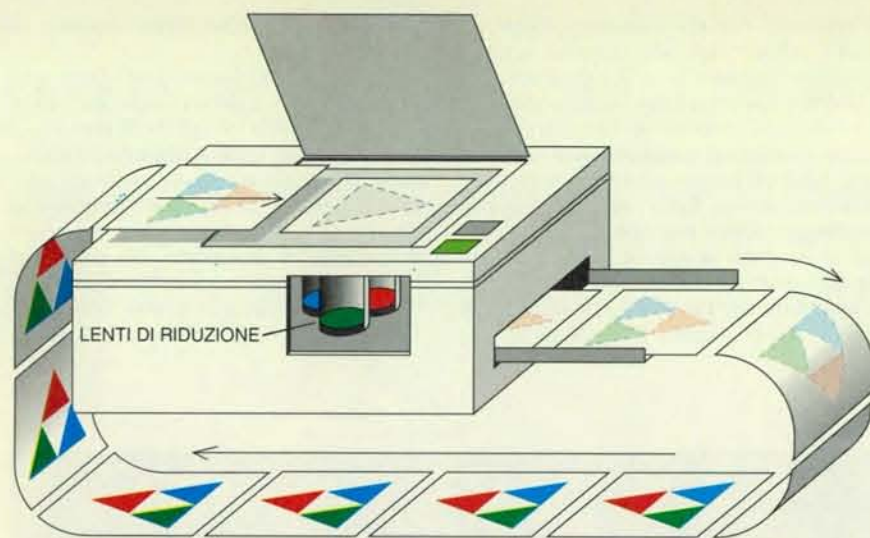
Certe lingue asiatiche come il cinese sono invece costituite da simboli che hanno di per sé un significato. Nel caso di queste lingue il numero di simboli o elementi possibili è arbitrariamente grande e si può considerare infinito. La geometria frattale è costruita più o meno allo stesso modo: è costituita da infiniti elementi, ciascuno dei quali è unico e completo. Gli elementi geometrici sono definiti da algoritmi, che hanno la funzione di unità «semantiche» della lingua frattale.

Le lingue frattali si dividono in due gruppi linguistici principali: quello lineare e quello non lineare. Le lingue di entrambi i gruppi, che «si parlano» usando un numero infinito di algoritmi, contengono un numero infinito di possibili immagini frattali; la lingua dei frattali non lineari, tuttavia, è molto più ricca e varia. In generale i dialetti seguono un insieme deterministico di regole (analoghe alle regole dell'ortografia e della grammatica). Vi è poi una famiglia, quella dei frattali aleatori, che differisce in quanto viene costruita senza regole deterministiche.

Il dialetto fondamentale della lingua frattale è la geometria frattale lineare. Questi frattali sono detti lineari perché i loro algoritmi hanno la stessa forma di quelli che definiscono le rette su un piano (in gergo matematico si dice che contengono solo termini del primo ordine).

Gli algoritmi lineari possono essere studiati con l'ausilio di un immaginario duplicatore di figure: la fotocopiatrice a riduzioni multiple (*si veda l'illustrazione nella pagina seguente in alto*). Si tratta di una metafora del magnifico lavoro di John E. Hutchinson, un matematico della Australian National University a Canberra. Questa macchina funziona più o meno come una normale fotocopiatrice con variatore di riduzione, ma ne differisce per il fatto di avere più lenti di riduzione, ciascuna delle quali può copiare l'originale collocato sulla macchina. Le lenti possono essere predisposte secondo diversi fattori di riduzione e le immagini ridotte possono essere collocate in qualsiasi posizione. La figura può quindi essere spostata, allungata, accorciata, riflessa, ruotata o trasformata in tutti i modi, purché i segmenti di retta dell'originale rimangano segmenti di retta.

Il modo in cui l'immagine viene spostata e ridotta è determinato dall'algoritmo. Mediante un anello di retroazione l'immagine viene elaborata ripetutamente, e tende via via a una forma frattale. Un esempio di frattale ottenuto con un algoritmo ricorsivo (a retroazione) è il triangolo di Sierpinski, dal nome del matematico polacco Waclaw Sierpinski che lo descrisse per primo nel 1916. Questo triangolo è autosimile: ciascuna

Questa immagine al calcolatore è una rappresentazione a tre dimensioni del potenziale elettrostatico che circonda un insieme di Mandelbrot dotato di carica. I valori di potenziale creano un paesaggio fantastico di catene montuose che circondano un vasto lago.

La macchina fotocopiatrice a riduzioni multiple genera una forma frattale tramite un anello a retroazione. Diverse lenti trasformano una immagine iniziale arbitraria (ingresso) in una nuova immagine (uscita), che è un *collage* di copie ridotte dell'immagine iniziale. L'immagine di uscita è poi passata più volte nella macchina dando origine all'immagine finale.

sua parte, per quanto piccola, contiene un'immagine che, ingrandita, riproduce tutto il triangolo.

Il triangolo di Sierpinski può essere costruito con una fotocopiatrice a riduzioni multiple nel modo seguente. Si pone nella macchina un'immagine, la si riduce della metà e la si copia tre volte, una su ciascun vertice di un triangolo equilatero. Il risultato è una configurazione triadica. Poi il procedimento si ripete: questa immagine triadica viene a sua volta ridotta della metà e copiata tre volte, e così via. Dopo che sono state compiute solo sei iterazioni comincia già a delinearsi una forma finale. Questa forma viene chiamata figura limite perché è il risultato al limite di un numero infinito di cicli della fotocopiatrice. La figura limite può essere approssimata in un tempo molto breve, ma non può essere mai ottenuta compiutamente.

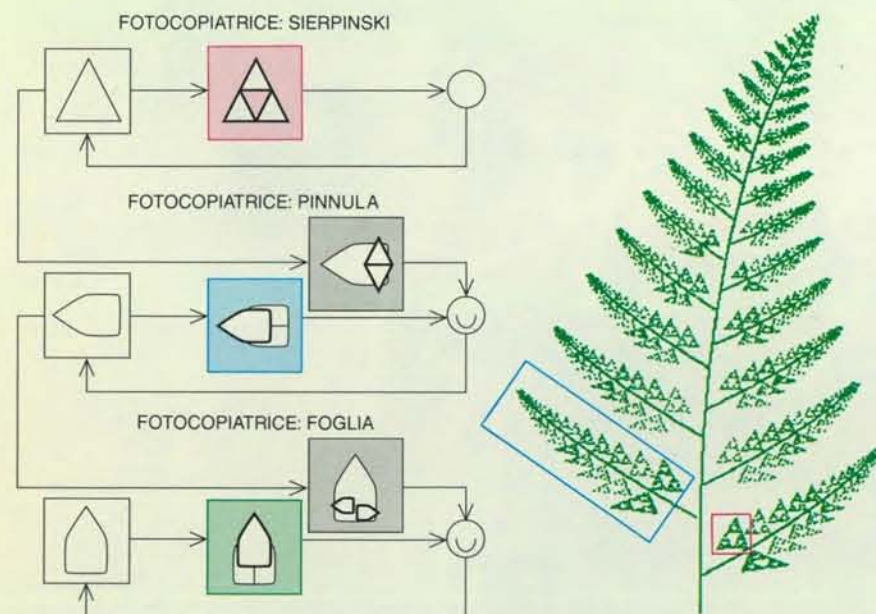
La figura limite non dipende dall'immagine di partenza. Anche se nella fotocopiatrice si colloca un'immagine di partenza molto particolare, come la parola FRACTAL, dopo sei cicli di copiatura la figura di partenza è quasi invisibile ed è la forma del triangolo di Sierpinski a predominare. A ogni copiatura le tracce dell'iniziale parola FRACTAL diventano sempre più indistinte.

Lievi ritocchi alla fotocopiatrice possono dar luogo a figure limite affatto diverse: un albero frattale o un frattale a forma di foglia di felce (*si veda l'illustrazione nella pagina a fronte*). La figura limite dipende solo dalle regole di riduzione e di spostamento (algoritmi) programmate nella macchina.

Queste regole sono casi particolari di quelle che i matematici chiamano trasformazioni lineari affini del piano, vale a dire trasformazioni che conservano la rettilinearità dei segmenti ma ne alterano la scala, la posizione e l'orientazione complessiva. Le regole di un dialetto frattale lineare possono essere descritte completamente da un certo numero (n) di funzioni di trasformazione, indicate con $[f_1, f_2, ..., f_n]$ (*si veda la parte superiore dell'illustrazione nella pagina a fronte*).

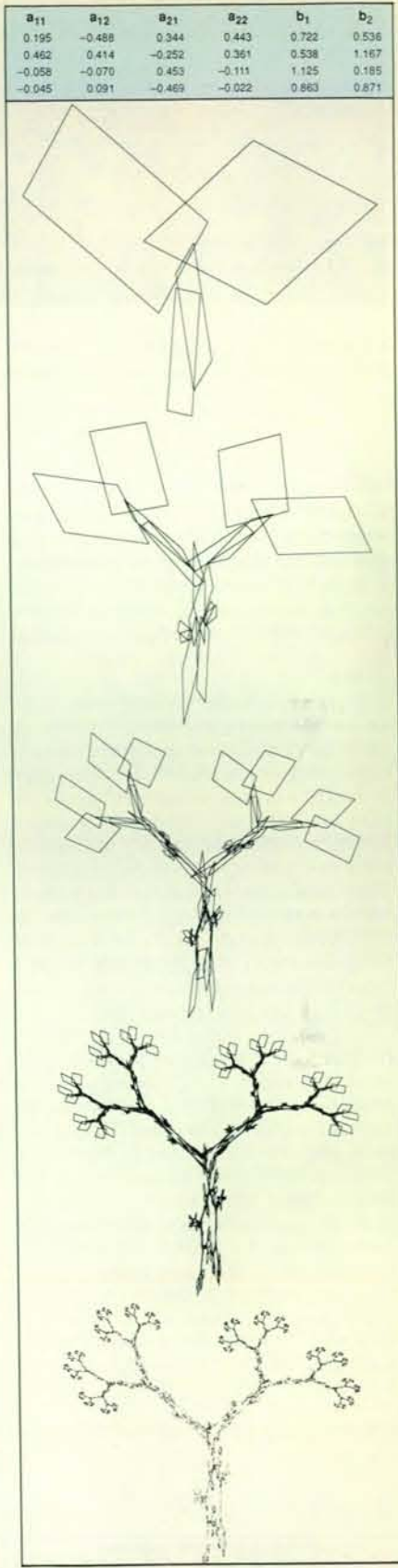
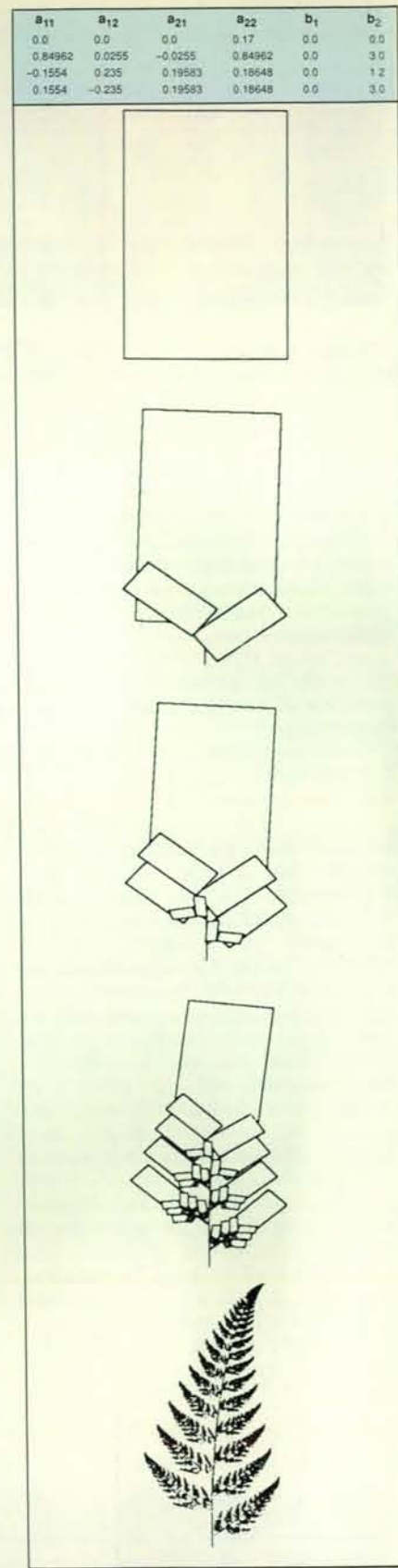
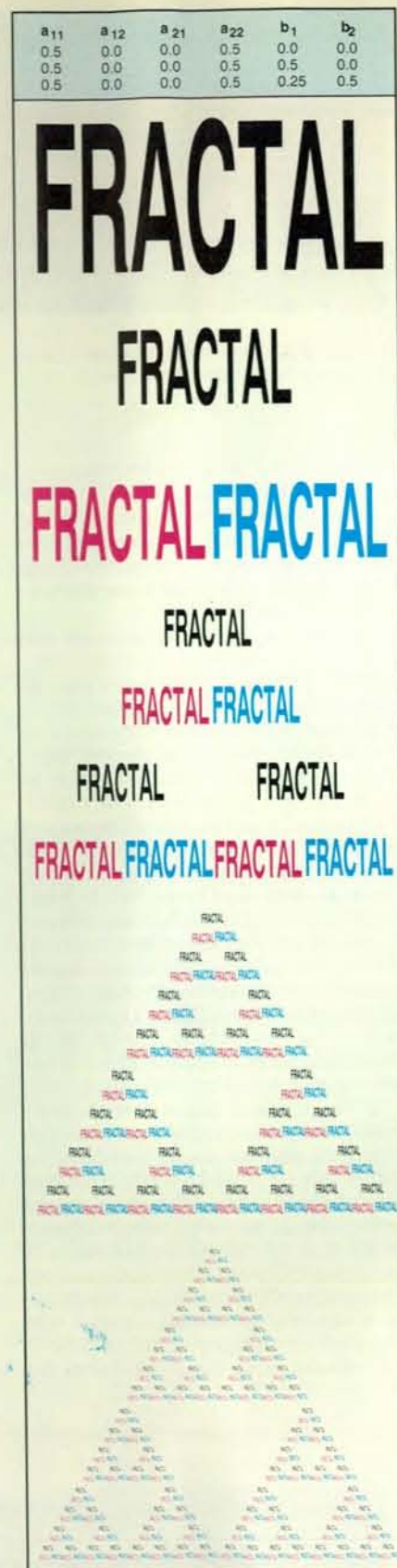
Questo fatto rappresenta una delle grandi potenzialità pratiche della geometria frattale. Descrivendo oggetti opportuni mediante un dialetto frattale lineare si può ridurre notevolmente la quantità di dati necessari per trasmettere o immagazzinare un'immagine. Una dimostrazione convincente di ciò è offerta da una foglia di felce. Una forma così complessa può essere descritta compiutamente da un algoritmo lineare basato su 24 numeri soltanto! Viceversa per rappresentare l'immagine della foglia punto per punto con la qualità di un'immagine televisiva ci vorrebbero parecchie centinaia di migliaia di valori numerici. In linea di principio qualsiasi immagine può essere codificata impiegando l'insieme opportuno di funzioni lineari di trasformazione.

Trasformando le immagini in codici mediante algoritmi frattali si potrebbe ridurre drasticamente il tempo, la complessità e il costo della loro trasmissione via satellite. Questa possibilità solleva un problema cruciale e in gran parte ancora non risolto. Come è possibile ottenere la più piccola famiglia di funzioni di trasformazione $[f_1, ... f_n]$ necessaria a definire un'immagine con una data precisione? Questo problema è attualmente oggetto di molti studi. Tra le applicazioni più generali di questo procedimento si potrebbero menzionare i codici per la costruzione di immagini a mezzatinta o perfino a colori.

La codificazione delle immagini frattali è utile solo se esiste qualche metodo efficiente per estrarre l'immagine imprigionata negli algoritmi frattali. La felce frattale ci fornisce un'opportunità molto interessante di capire come viene prodotta l'immagine. Le regole da fornire

Più macchine fotocopiatrici collegate in una rete possono creare immagini frattali composite, come questa foglia di felce costituita da triangoli di Sierpinski. Si collegano in parallelo più macchine: la prima genera i triangoli di Sierpinski, la seconda li dispone in pinnule e la terza genera la forma totale della foglia (*a sinistra*). Si noti che le pinnule si alternano a destra e a sinistra dello stelo principale; su di esse i triangoli sono contrapposti (*a destra*).

a_{11}	a_{12}	a_{21}	a_{22}	b_1	b_2
0.5	0.0	0.0	0.5	0.0	0.0
0.5	0.0	0.0	0.5	0.5	0.0
0.5	0.0	0.0	0.5	0.25	0.5

a_{11}	a_{12}	a_{21}	a_{22}	b_1	b_2
0.0	0.0	0.0	0.17	0.0	0.0
0.84962	0.0255	-0.0255	0.84962	0.0	3.0
-0.1554	0.235	0.19583	0.18648	0.0	1.2
0.1554	-0.235	0.19583	0.18648	0.0	3.0

a_{11}	a_{12}	a_{21}	a_{22}	b_1	b_2
0.195	-0.488	0.344	0.443	0.722	0.536
0.462	0.414	-0.252	0.361	0.538	1.167
-0.058	-0.070	0.453	-0.111	1.125	0.185
-0.045	0.091	-0.469	-0.022	0.863	0.871

Le immagini frattali generate dall'anello di retroazione della macchina fotocopiatrice dipendono solo dal suo programma di copiatura. La parola FRACTAL viene trasformata da un programma che riduce a metà la grandezza di un'immagine e la copia tre volte, una per ogni vertice di un triangolo equilatero. L'immagine risultante è un triangolo di Sierpinski (*a sinistra*). Trasformazioni analoghe ma un poco più elaborate danno origine a un frattale felciforme (*al centro*) o a un albero frattale (*a destra*). Qualsiasi immagine iniziale introdotta nella fotocopiatrice fornirebbe lo stesso risultato. Sono sufficienti i pochi numeri che definiscono le regole di copiatura (*in alto*) per specificare un'immagine che per essere descritta con metodi tradizionali richiederebbe centinaia di migliaia di numeri.

Gli insiemi di Julia sono frontiere frattali che vengono generate dall'iterazione della trasformazione quadratica $z^2 + c$. La straordinaria varietà delle loro forme dipende solo dal valore del parametro di controllo c. Se il punto c appartiene all'insieme di Mandelbrot, l'insieme di Julia corrispondente è connesso (*nella pagina a fronte*). Gli insiemi di Julia definiti da punti non appartenenti all'insieme di Mandelbrot non sono connessi (*in questa pagina*). L'insieme di Mandelbrot funge da indice degli insiemi di Julia.

alla fotocopiatrice per ottenere questo frattale stabiliscono che ciascuna trasformazione dia luogo a quattro riduzioni e spostamenti dell'immagine precedente. Una delle trasformazioni attua una riduzione molto rapida che schiaccia l'immagine, riducendola a un segmento verticale; e questo segmento costituisce lo «stelo».

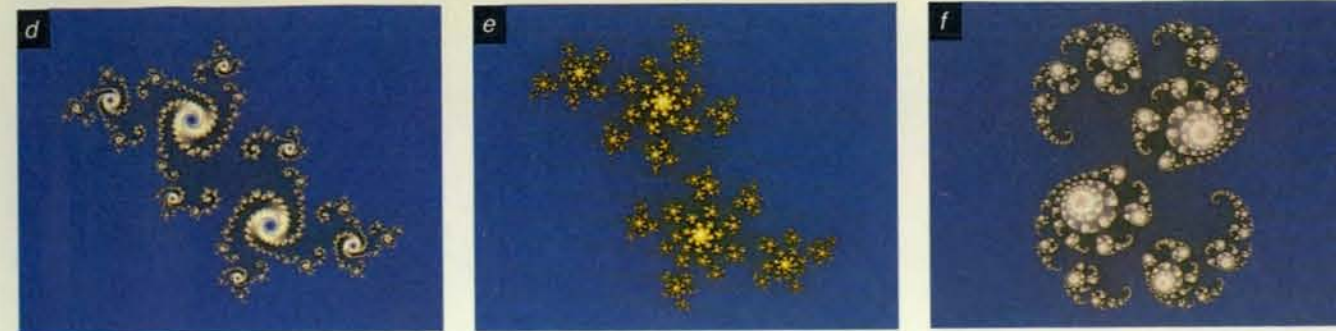
Se si comincia con un solo rettangolo, a ogni copiatura il numero dei rettangoli aumenta di un fattore quattro, e dopo m trasformazioni è pari a 4^m. Dopo quattro iterazioni l'immagine iniziale (in questo caso un rettangolo) può essere ancora facilmente identificata. Un rettangolo abbastanza piccolo da rendere visibile la figura limite (la foglia di felce) si potrebbe ottenere dopo circa 50 iterazioni. Si dovrebbero quindi calcolare e disegnare 4^{50} (più o meno 10^{30}) rettangoli. Nessun calcolatore esistente sarebbe all'altezza di un compito così impegnativo.

Vista questa difficoltà, viene da domandarsi come si possano produrre queste figure limite. L'espediente che consente di ottenere queste immagini è un algoritmo che chiameremo il gioco del caos, proposto da Michael E. Barnsley e Stephen Demko del Georgia Institute of Technology. Il gioco ha inizio con la scelta di un punto arbitrario del piano. Poi viene lanciato un dado a quattro facce, ciascuna delle quali corrisponde a una delle quattro trasformazioni che generano la figura della foglia di felce. Lanciando il dado si individua a caso una delle trasformazioni $[f_1, f_2, f_3, f_4]$, che viene poi applicata al punto scelto e lo trasforma in un nuovo punto del piano. Un altro lancio individua un'altra trasformazione, che viene applicata al punto ottenuto in precedenza e così via. I punti generati dai lanci successivi convergono rapidamente e riempiono densamente la figura limite. Il problema di questa tecnica è che per ottenere la figura limite potrebbe essere necessario un tempo lunghissimo.

Nell'esempio precedente tutte le f_k (dove k rappresenta semplicemente una delle funzioni possibili) hanno la stessa probabilità di essere scelte mediante il lancio del dado. La figura limite può essere ottenuta molto più rapidamente se le probabilità P_k con cui vengono scelte le funzioni f_k nel gioco del caos non sono tutte uguali, cioè se alcune f_k sono più probabili di altre. Il modo più rapido per ottenere la figura limite è quello di assegnare la probabilità più elevata alle funzioni che effettuano la minima riduzione dell'immagine. Con questa modificazione ciascun punto della figura limite è individuato con la stessa frequenza nel gioco del caos, quindi tutte le parti della figura si riempiono con la stessa velocità.

Modificando il gioco del caos si possono descrivere le mezzetinte semplicemente traducendo la frequenza con cui un punto della figura è individuato in un valore della scala dei grigi. Con una scelta opportuna delle P_k, per ogni punto della figura si può ottenere il valore desiderato della scala dei grigi (cioè la frequenza desiderata con cui il punto viene individuato). Applicando questa tecnica ai colori primari additivi (rosso, verde e blu) si possono codificare immagini a colori. In questo modo l'utilità della compressione frattale dei dati viene ulteriormente migliorata.

Per ora non esiste alcun metodo soddisfacente per generare in modo automatico codifiche frattali di una data figura o immagine. Per immagini autosimili come la felce di Barnsley esiste un procedimento semiautomatico che comporta un'interazione tra calcolatore e osservatore. Per prima cosa si scompone l'immagine in parti simili all'immagine complessiva. Nel caso della foglia di felce le due pinnule inferiori sono di forma simile al tutto, così come lo è la parte superiore della foglia che resta quando le pinnule inferiori sono soppresse. Consideriamo una fotocopiatrice a riduzioni multiple contenente trasformazioni che riducano l'immagine intera a queste parti. Ciò si può ottenere facilmente con un procedimento per tentativi ed errori basato su un programma interattivo.

L'idea di base di questo metodo comporta che solo immagini rigorosamente autosimili possano essere codificate in forma frattale. Questa limitazione può essere superata grazie a una promettente estensione del metodo, che è attualmente allo studio. L'idea di fondo è quella di avere molte fotocopiatrici funzionanti simultaneamente in parallelo e organizzate in una rete gerarchica. Questa rete può controllare singoli caratteri autosimili oppure mescolarne più d'uno; ciò consente, per esempio, di creare una foglia felciforme composta di triangoli di Sierpinski (*si veda l'illustrazione a pagina 44 in basso*).

Occupiamoci ora di un altro insieme di dialetti frattali, i dialetti non lineari. Uno di essi, il dialetto quadratico, è stato oggetto di attenzione particolare, poiché produce una grande ricchezza di forme geometriche a partire da un algoritmo piuttosto semplice ed è strettamente collegato all'odierna teoria del caos.

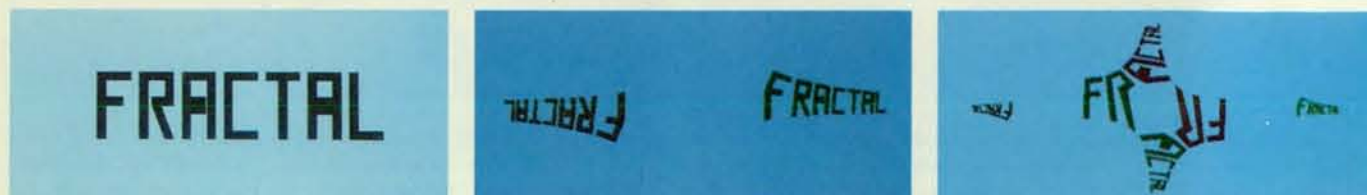
La teoria su cui si basa il dialetto quadratico fu descritta per la prima volta nel 1918 dal matematico francese Gaston Julia, che si trovava allora in un ospedale militare, convalescente delle ferite riportate durante la prima guerra mondiale. Tanto le sue ricerche quanto quelle contemporanee del suo accanito rivale Pierre Fatou furono presto quasi dimenticate, ma di recente il lavoro di Mandelbrot ha riacceso l'interesse per le loro teorie. L'impresa intellettuale di Julia e Fatou è particolarmente notevole perché, non esistendo a quel tempo i calcolatori, essi potevano contare solamente sulle proprie capacità intrinseche di visualizzazione.

Julia e Fatou si occupavano di numeri complessi, ciascuno costituito da un numero reale e da un multiplo di i, l'unità immaginaria definita come la radice quadrata di -1. I numeri complessi vengono di solito rappresentati su un piano con due assi perpendicolari, uno dei quali corrisponde ai numeri reali e l'altro ai numeri immaginari. I due studiosi cercavano di capire che cosa accade a una successione di punti z_k del piano dei numeri complessi generati dalla trasformazione $g(z) = z^2 + c$. Il punto z_{k+1} si ottiene applicando la trasformazione al punto precedente della successione z_k. Il numero complesso c è un parametro di controllo che può essere scelto ad arbitrio. Questo processo iterativo, in apparenza semplice, costituisce la base di una famiglia sbalorditiva di forme.

Quando si applica la trasformazione a un punto iniziale z_0, la successione risultante può comportarsi in due modi diversi: può vagare senza limitazioni, allontanandosi verso l'infinito, oppure restare confinata in una certa regione del piano dei numeri complessi. I punti liberi costituiscono il cosiddetto insieme di fuga; quelli che restano confinati formano il cosiddetto insieme prigioniero. Se il punto di partenza z_0 appartiene all'insieme prigioniero, esso genera una successione che resta in una «prigione» numerica indipendentemente da quante generazioni della successione siano calcolate. La forma della prigione dipende dal valore di c scelto. Per un punto z_0 esterno all'insieme prigioniero, la successione z_k si allontana dal centro del piano e va verso l'infinito. L'insieme prigioniero e l'insieme di fuga sono separati da una frontiera infinitamente stretta, il cosiddetto insieme di Julia (*si veda l'illustrazione in alto in queste due pagine*).

Per quanto appaia sorprendente, l'insieme di Julia si può ottenere anche usando la fotocopiatrice a riduzioni multiple, equipaggiandola con lenti speciali che invertano l'effetto di $g(z)$. L'inversione di $g(z) = z^2 + c$ si compie con due funzioni di trasformazione, $f_1(u) = + (u - c)^{1/2}$ e $f_2(u) = - (u - c)^{1/2}$. (In queste funzioni c è il solito parametro di controllo e u è il valore iniziale scelto.) Queste due funzioni possono essere considerate le «riduzioni» effettuate dalla fotocopiatrice. Facendo funzionare la macchina ripetutamente, punti scelti a caso tendono all'insieme di Julia.

La presenza della radice quadrata nelle equazioni significa che la fotocopiatrice non applica più un fattore di riduzione uniforme. Inoltre, poiché la trasformazione è non lineare, i segmenti di retta vengono trasformati in linee curve. Da un'immagine iniziale emergono due immagini più piccole, poi quattro, poi otto, finché pian piano si forma la figura limite (*si veda l'illustrazione in basso in queste due pagine*). Come nel caso dei frattali lineari, la figura limite non dipende dalla particolare immagine di partenza, ma è completamente determinata da f_1 e f_2, cioè dalla scelta del parametro c.

Si presenta ora uno dei problemi più difficili e affascinanti della geometria frattale. Tornando alla metafora della lingua, il problema può essere tradotto in questa domanda: quali sono le regole grammaticali dei dialetti quadratici? In termini matematici il problema è: l'infinita varietà degli insiemi di Julia ammette un principio ordinatore soggiacente?

Anche i frattali non lineari come gli insiemi di Julia possono essere generati tramite una fotocopiatrice a riduzioni multiple opportunamente programmata. La macchina ora non effettua più semplici operazioni di riduzione, ma curva e distorce l'immagine di partenza oltre a rimpicciolirla. La fotocopiatrice compie per via grafica un'inversione delle equazioni quadratiche che definiscono un insieme di Julia effettuando le trasformazioni $+(z - c)^{1/2}$ e $-(z - c)^{1/2}$. La figura limite generata dalla fotocopiatrice è un insieme di Julia.

Nel cercare la risposta si è giunti a una delle più belle scoperte della matematica sperimentale. La soluzione sta nel fatto, noto a Julia e a Fatou, che per ogni parametro di controllo c l'immagine frattale risultante è di due tipi possibili: l'insieme di Julia può essere un unico insieme connesso, oppure essere costituito da un numero infinito di punti non connessi, come polvere.

Supponiamo di tracciare un punto per ogni parametro di controllo c del piano complesso appartenente a un insieme di Julia connesso e di lasciare uno spazio vuoto per ogni c appartenente a un insieme di Julia non connesso. Il risultato è l'insieme, ormai celebre, di Mandelbrot, un frattale dalla ricchezza davvero straordinaria.

Naturalmente per decidere se un punto c appartiene all'insieme di Mandelbrot, è necessario sapere se un dato insieme di Julia è connesso. Uno dei grandi successi di Julia e Fatou fu la scoperta che questo difficile problema può essere risolto con un semplice calcolo. Si consideri la successione di valori z_k generata dalla funzione $g(z) = z^2 + c$ quando il punto iniziale z_0 coincide con lo zero. In tal modo l'attenzione si concentra sul fattore cruciale, il parametro di controllo c. La successione risultante è 0, c, $c^2 + c$, $(c^2 + c)^2 + c$, ... Se questa successione non diverge verso l'infinito allora il corrispondente insieme di Julia è connesso e il punto c appartiene all'insieme di Mandelbrot.

Ogni porzione dell'insieme di Mandelbrot caratterizza una famiglia di insiemi di Julia correlati. Per esempio il corpo principale, a cardioide, dell'insieme di Mandelbrot caratterizza gli insiemi di Julia che somigliano a cerchi raggrinziti. Benché l'insieme di Mandelbrot non sia esattamente autosimile come il triangolo di Sierpinski e la foglia di felce frattale, esso possiede una proprietà analoga: se si ingrandisce la sua frontiera, l'insieme di Mandelbrot rivela un numero infinito di minuscole copie di se stesso. La ricchezza di forme e di strutture contenute nell'insieme di Mandelbrot può essere apprezzata solo quando lo si esamini con grande minuzia.

La proprietà forse più affascinante dell'insieme di Mandelbrot è che esso può essere considerato un «deposito» di immagini di efficienza infinita: oltre a suddividere gli insiemi di Julia in connessi e non connessi, l'insieme di Mandelbrot funge anche da indice diretto e grafico di un numero infinito di insiemi di Julia. Ingrandendo l'insieme di Mandelbrot intorno a un punto c situato sulla sua frontiera, appaiono forme che sono anche gli elementi costitutivi dell'insieme di Julia corrispondente al punto c. Questa scoperta, tuttavia, non è stata ancora rivestita di tutto il necessario rigore matematico. Tan Lei, un giovane ricercatore di talento che lavora all'Università di Lione, ha dimostrato che l'insieme di Mandelbrot si comporta in questo modo per la maggior parte dei valori del parametro c situati esattamente sulla frontiera dell'insieme.

Le proprietà dell'insieme di Mandelbrot sono state e sono tuttora un gran de cimento per la ricerca matematica. Enormi progressi sono stati compiuti grazie alla fusione di teoria matematica ed esperimenti di eidomatica (ossia di grafica al calcolatore), in particolare nei fondamentali lavori di Adrien Douady dell'École Normale Supérieure di Parigi e di John H. Hubbard della Cornell University.

Il lavoro di gran lunga più riuscito in questo campo è quello sul cosiddetto potenziale elettrostatico dell'insieme di Mandelbrot. Si immagini che l'insieme sia dotato di carica elettrica. Si potrebbe misurare il potenziale collocando una carica puntiforme all'esterno dell'insieme e misurando la forza elettrostatica agente su quel punto. Risulta che il calcolo del potenziale è strettamente legato alla serie 0, c, $c^2 + c$, $(c^2 + c)^2 + c$..., usata per stabilire se un punto c appartiene o no all'insieme di Mandelbrot.

Si è visto che generare una rappresentazione tridimensionale del potenziale è laborioso, specie nelle animazioni usate per studiare l'insieme di Mandelbrot. Esaminando meglio le proprietà eidomatiche del potenziale, di recente si è potuto ridurre di un ordine di grandezza il costo in tempo macchina. Di conseguenza molti ricercatori, noi compresi, studiano sempre più l'insieme di Mandelbrot mediante video animati dal calcolatore. Ricerche analoghe sono in corso anche su rappresentazioni tridimensionali del potenziale di altri frattali.

Tutti i frattali finora esaminati possono essere considerati deterministici. Benché i processi aleatori (per esempio il lancio di un dado) possano aiutarci a produrre immagini frattali, essi non hanno alcun effetto sulla forma frattale finale. La situazione è affatto diversa per un'altra classe di frattali, i cosiddetti frattali aleatori.

Per generare un frattale di questo tipo si può cominciare con un triangolo giacente su un piano arbitrario. I punti medi di ciascun lato del triangolo vengono collegati tra loro e il triangolo viene così diviso in quattro triangoli più piccoli. Ciascun punto medio viene poi alzato o abbassato di una quantità scelta a caso. Lo stesso procedimento viene applicato a ciascuno dei triangoli più piccoli e il processo viene ripetuto all'infinito. All'aumentare del numero delle iterazioni, comincia a formarsi una superficie sempre più ricca di particolari.

In questo «metodo dello spostamento dei punti medi», l'entità aleatoria dello spostamento dei punti medi è retta da una legge di distribuzione che può essere modificata fino a ottenere una buona approssimazione della superficie di cui si vuol costruire il modello. Per un modello di una superficie relativamente liscia, le trasformazioni usate dovrebbero prevedere una regola per cui gli spostamenti dei punti medi diventino piccolissimi già dopo poche iterazioni. Una regola del genere aggiunge solo piccole prominenze sullo sviluppo complessivo. Per rappresentare invece una superficie accidentata, per esempio la topografia di una catena montuosa, è meglio far diminuire di poco l'entità degli spostamenti a ogni iterazione.

Si possono creare paesaggi frattali con il metodo dello spostamento dei punti medi. I punti medi dei lati di un triangolo (*a*) vengono uniti da segmenti e spostati in su o in giù, fuori dal piano dell'immagine (*b*). Si ottengono così quattro piccoli triangoli su cui si ripete il procedimento. Una legge di distribuzione stabilisce l'entità dello spostamento e quindi determina la scabrosità del terreno frattale. Un programma eidomatico genera poi ombreggiature appropriate (*c*), dando vita a risultati straordinariamente realistici (*d*).

L'insieme di Mandelbrot rispecchia l'ordine soggiacente alla infinita varietà degli insiemi di Julia. Tutti i suoi punti rappresentano valori del parametro c corrispondenti a insiemi di Julia connessi. Se il punto c non appartiene all'insieme di Mandelbrot l'insieme di Julia a esso associato non è connesso. L'insieme di Mandelbrot contiene una ricchezza di dettagli inimmaginabile. Tre ingrandimenti successivi dell'insieme rivelano strutture simili che si ripetono, fra cui anche copie in miniatura dell'insieme stesso, oltre a forme nuove e differenti. Se l'intero insieme fosse raffigurato alla scala dell'immagine di destra, sarebbe grande come 100 campi da calcio.

Questo metodo per costruire superfici ha molte applicazioni. È stato impiegato per ottenere modelli dell'erosione del suolo e per analizzare le registrazioni sismiche al fine di capire i cambiamenti nelle zone di faglia. Questo concetto è stato usato da Richard F. Voss, collega di Mandelbrot al Thomas J. Watson Research Center, per generare immagini molto realistiche di pianeti, satelliti, nubi e montagne (*si veda l'illustrazione nella pagina a fronte*).

A prescindere dalla loro origine e dal metodo di costruzione, tutti i frattali presentano una caratteristica importante: se ne può misurare la scabrosità, la complessità o l'accartocciamento mediante un numero caratteristico, la dimensione frattale. Le varie definizioni concettuali della dimensione frattale risalgono più o meno a un lavoro del 1919 del matematico Felix Hausdorff dell'Università di Bonn.

La dimensione frattale può essere determinata mediante un procedimento di conteggio proposto da Mandelbrot. Si consideri una forma complessa a cui è sovrapposto un reticolo di quadrati tracciati su carta millimetrata. Alcuni quadrati conterranno parte della forma, altri saranno vuoti. Il numero N di quadrati non vuoti dipende dalla forma data e dal lato E dei quadrati del reticolo. Si postula che N sia proporzionale a $1/E^D$ (più è fitto il reticolo più sono i quadrati non vuoti). L'esponente D è la dimensione. Per una figura piana, per esempio un cerchio, quando il lato dei quadrati del reticolo viene ridotto a metà il numero dei quadrati non vuoti dovrebbe essere moltiplicato per quattro (due al quadrato), perché la figura ha dimensione due. Per un frattale il numero dei quadrati non vuoti verrebbe moltiplicato per un valore frazionario un po' più grande o un po' più piccolo.

Questo procedimento non è limitato alle forme o agli oggetti matematici contenuti in un piano: si può anche calcolare la dimensione frattale di cose reali come fiumi, nubi, litorali, alberi, arterie o villi intestinali. Le arterie umane, per esempio, hanno una dimensione frattale di circa 2,7.

Oltre a essere utile per descrivere la complessità degli oggetti naturali, la geometria frattale offre un'interessante possibilità per rinnovare l'insegnamento della matematica. I concetti della geometria frattale sono evidenti e intuitivi e le forme che s'incontrano possiedono una grande attrattiva estetica e un'ampia gamma di applicazioni. La geometria frattale può quindi contribuire a sfatare l'idea che la matematica sia arida e inaccessibile e può motivare gli studenti ad apprendere questa sconcertante ed eccitante disciplina.

Di fronte al linguaggio dei frattali, così nuovo e in evoluzione tanto rapida, gli stessi scienziati e matematici provano una sorta di stupore infantile. Scrive Mandelbrot:

«Gli scienziati... saranno sorpresi e compiaciuti nello scoprire che non poche forme che prima dovevano chiamare *granulose*, *tentacolari*, *intermedie*, *foruncolose*, *butterate*, *ramificate*, *algali*, *strane*, *intricate*, *tortuose*, *serpeggianti*, *esili*, *grinzose* e così via potranno d'ora in poi essere studiate in modo rigorosamente e vigorosamente quantitativo.

«I matematici... saranno sorpresi e compiaciuti nello scoprire che gli insiemi [frattali] considerati finora eccezionali... costituiscono in un certo senso la regola, che certe costruzioni ritenute patologiche discendono in modo naturale da problemi molto concreti e che lo studio della natura può contribuire a risolvere problemi antichi e a trovarne tanti di nuovi.»

BIBLIOGRAFIA

MANDELBROT B. B., *The Fractal Geometry of Nature*, W. H. Freeman & Co., 1983.

PEITGEN H. O. e RICHTER P., *La bellezza dei frattali*, Bollati Boringhieri, Torino, 1987.

BARNSLEY M., *Fractals Everywhere*, Academic Press, Inc., 1988.

PEITGEN H. O. e SAUPE D., *The Science of Fractal Images*, Springer-Verlag, 1988.

PEITGEN HEINZ-OTTO, JÜRGENS HARTMUT e SAUPE DIETMAR, *Fractals for the Classroom*, Springer-Verlag, 1989.

PEITGEN H. O., JÜRGENS H., SAUPE D. e ZAHLTEN C., *Fractals: An Animated Discussion*, con Edward Lorenz e Benoit B. Mandelbrot (video), W. H. Freeman & Co., 1990.

La singolare termodinamica dei microaggregati

Piccoli gruppi costituiti da alcune decine di atomi o molecole, a causa dei loro punti di fusione e di congelamento che non coincidono, presentano proprietà sorprendenti come la coesistenza delle fasi liquida e solida

di R. Stephen Berry

Quasi tutti i solidi hanno un punto di fusione e quasi tutti i liquidi hanno un punto di congelamento. In realtà questi due punti sono esattamente identici, ma visti da due prospettive diverse: il ghiaccio fonde a zero gradi Celsius - la temperatura più elevata alla quale può esistere come solido stabile - mentre l'acqua gela a zero gradi Celsius, ossia la temperatura più bassa alla quale può esistere come liquido stabile. Sembra che non vi sia nulla di più semplice.

Tuttavia, piccoli aggregati di atomi o molecole consentono di osservare il punto di fusione e quello di congelamento sotto una nuova luce, e soprattutto mostrano come non siano così semplici da definire. Questi microaggregati (composti da un numero di atomi o molecole che può andare dai quattro-cinque ai 100-200) possono coesistere allo stato di solido e di liquido in un intervallo finito (ossia non nullo) di temperature e hanno punti di fusione e di congelamento notevolmente diversi.

Sono state le peculiari caratteristiche dei microaggregati a permettere a molti ricercatori, me compreso, di studiare i segreti dei loro punti di fusione e di congelamento. Questi aggregati sono più grandi di singole molecole, ma più piccoli di una massa ordinaria di materia (che è composta da un numero di atomi così elevato da poter essere considerato infinito) e quindi presentano le proprietà di entrambe. Grazie alla loro dimensione intermedia, essi possono essere studiati quasi con la stessa minuziosa precisione degli atomi o delle molecole, ma nello stesso tempo mostrano alcune delle proprietà globali della materia.

In futuro gli aggregati atomici potranno anche offrire la possibilità di produrre nuovi tipi di materiali e di dar luogo a nuove reazioni chimiche. Dato che gli aggregati possono esistere in molte forme stabili differenti - per esempio come icosaedri da 55 atomi o come «palloni da calcio» da 60 atomi - potrebbero essere fatti condensare in solidi diversi da tutti quelli finora conosciuti. Gli studiosi sperano che sia possibile progettare questi materiali in modo che abbiano le proprietà microelettroniche, meccaniche o catalitiche desiderate.

Dall'epoca di John Dalton (inizio del XIX secolo), quando la teoria atomica cominciò a essere accettata, lo studio del comportamento della materia si è diviso in due branche. I «riduzionisti» si sono concentrati sulle proprietà dei singoli atomi e molecole, e negli anni trenta del nostro secolo questo filone di ricerca ha dato vita alla fisica nucleare e poi alla fisica delle particelle. Altri ricercatori si sono occupati delle proprietà globali della materia, soprattutto dei grandi agglomerati di atomi o molecole. I microaggregati permettono di collegare queste due aree di ricerca, ma per poterli studiare si dovette attendere lo sviluppo di adeguate metodologie sperimentali e teoriche, che divennero disponibili solo all'inizio degli anni settanta.

Al contrario delle molecole, che sono caratterizzate da una composizione e, nella maggior parte dei casi, da una struttura ben determinata, i microaggregati non sono definiti in maniera altrettanto precisa. Un microaggregato di atomi di silicio, per esempio, può contenere tre, o 10 o 100 atomi. Oltre a ciò, la maggior parte di questi aggregati può presentare un gran numero di strutture, tutte stabili anche se in misura diversa.

Gli aggregati differiscono dalla materia ordinaria non solo per il numero di atomi o molecole che li costituiscono, ma anche per il numero di atomi o molecole che si trovano alla superficie. In un campione di materia ordinaria solo una piccola percentuale degli atomi presenti si trova alla superficie, mentre negli aggregati questa percentuale può essere molto più grande. In un aggregato di 55 atomi di argo, per esempio, almeno 42 atomi sono in superficie.

I principi teorici in base ai quali si può spiegare la differenza nei punti di fusione e di congelamento degli aggregati si fondano su diversi concetti, in particolare su quello di «buca di potenziale» che può essere assimilato all'esperienza quotidiana legata alla forza di gravità. Si consideri una regione fatta di colline e valli; il fondo di una valle - la buca - è, a causa dell'attrazione di gravità, il punto di minima energia potenziale per qualunque oggetto dotato di massa. Una palla posta sopra una collina rotola verso il basso, mentre una che si trova in fondo a una valle rimane ferma. Una forza attrae o spinge un oggetto nella direzione in cui si ha una riduzione della sua energia potenziale.

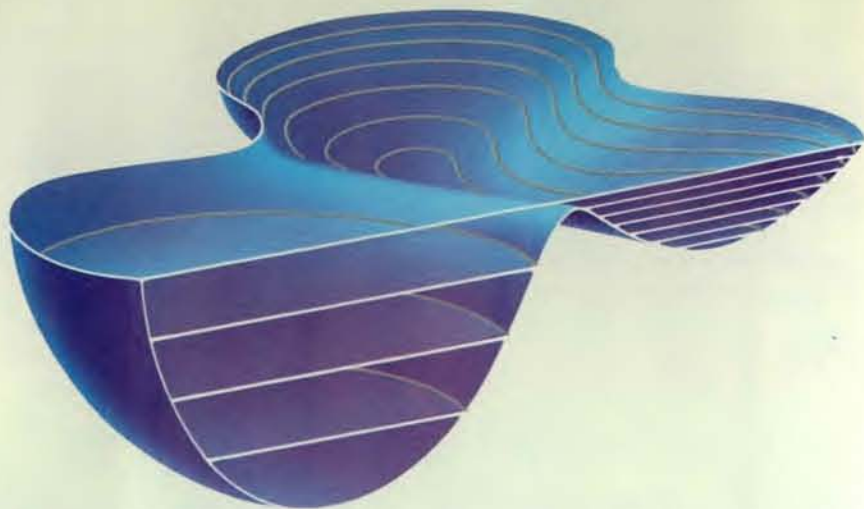
Si può immaginare che una molecola o un aggregato che si trovi esattamente all'equilibrio giaccia sul fondo di una buca di potenziale, ma in questo caso sono le forze elettriche, non quella di gravità, ad agire per una riduzione dell'energia potenziale. Quest'azione viene esercitata dalla forza attrattiva tra coppie di elettroni e di protoni degli atomi costituenti e dalla forza repulsiva all'interno delle coppie di protoni e di elettroni. Sono i contributi di queste due forze a determinare il diametro di un atomo, ossia, più esattamente, la distanza dal centro alla quale si può avere una collisione con un altro atomo. Gli atomi infatti non sono sfere dure come palle da biliardo, ma sono in una certa misura comprimibili e deformabili.

La buca di potenziale di due atomi che interagiscono per formare un legame ha un valore costante quando gli atomi sono lontani; tale valore, però, diminuisce quando essi si avvicinano, a causa dell'attrazione elettrone-protone, per poi aumentare bruscamente in corrispondenza della collisione fra gli atomi. Benché sia relativamente semplice estendere il concetto di buca di potenziale a un'astrazione che rappresenti le mutue interazioni fra tre o più atomi, è difficile rappresentare una buca di questo tipo perché ha un numero troppo grande di dimensioni. In altri termini, la buca dipende da troppe variabili indipendenti per poter essere raffigurata su carta o anche rappresentata con un modello tridimensionale. Tuttavia la terminologia e le immagini associate al concetto di buca di potenziale rimangono utili per descrivere il comportamento di molecole e aggregati formati da diversi atomi.

Finora gli aggregati sono stati descritti in termini di meccanica classica, ma un'interpretazione migliore della natura è fornita dalla meccanica quantistica. Il cambiamento più importante introdotto dalla descrizione quantistica è che le energie permesse di un aggregato sono quantizzate. Da un punto di vista classico, naturalmente, un aggregato può assumere qualsiasi energia nella buca di potenziale, mentre dal punto di vista quantistico l'energia dell'aggregato è limitata a una serie discreta di livelli, che formano una sorta di irregolare «scaletta» nella buca di potenziale.

Il fatto che i «pioli» - o livelli quantizzati di energia - di questa scaletta non siano distanziati in maniera uniforme è significativo nel determinare i differenti punti di fusione e di congelamento dei microaggregati: una buca di potenziale profonda e ripida ha livelli ampiamente spaziati, mentre una buca di potenziale larga li ha molto ravvicinati. Una buca può essere profonda e ripida nel centro e avere margini ampi e poco profondi; in questo caso i livelli di bassa energia sono molto distanziati e quelli di alta energia sono ravvicinati.

Gli aggregati in fase solida, che vibrano con difficoltà e possono ruotare in maniera limitata, come i modelli fatti di bastoncini e sferette, sono associati a buche profonde. Le buche poco pronunciate, più simili a semplici ondulazioni, sono correlate agli aggregati in fase liquida, che non hanno una struttura geometrica definita e quindi si riorganizzano facilmente. Gli aggregati in fase solida hanno livelli di energia relativamente distanziati, mentre quelli in fase liquida li

Un aggregato di 19 atomi di argo può esistere in fase solida (*in alto*) o in fase liquida (*in basso*). Un aggregato in fase solida può avere solo vibrazioni limitate mentre uno in fase liquida non ha una struttura geometrica definita e può quindi riorganizzarsi facilmente. Le immagini al calcolatore sono state realizzate da Thomas L. Beck, che oggi lavora presso l'Università di Cincinnati.

Il concetto di buca di potenziale è utile per descrivere il comportamento degli aggregati. Una molecola o un aggregato in condizioni di equilibrio può essere immaginata come sul fondo di una buca di potenziale, proprio come una palla che, in fondo a una valle, rimane ferma. Da un punto di vista classico un aggregato può assumere qualsiasi energia della buca di potenziale, mentre da un punto di vista quantistico i valori possibili dell'energia sono limitati a un gruppo di livelli discreti (*piani orizzontali*) che formano i «pioli» di una sorta di scaletta irregolare nella buca. Una buca di potenziale profonda e ripida ha livelli di energia molto distanziati ed è associata ad aggregati in fase solida, mentre una buca ampia e poco profonda, con livelli ravvicinati, è correlata ad aggregati in fase liquida.

hanno molto ravvicinati. Riprendendo l'analogia con un paesaggio collinoso, si può dire che, nella situazione in cui sono più probabili differenze fra la temperatura di fusione e quella di congelamento, le valli strette e profonde della buca di potenziale sono separate dalle ampie e ondulate pianure da passi moderatamente alti oppure inaccessibili.

L'ultimo concetto necessario per spiegare la differenza nei punti di fusione e di congelamento nei microaggregati è quello di energia libera: essa viene definita come la differenza fra l'energia interna del sistema - in questo caso un insieme di molecole - e il prodotto della sua temperatura per la sua entropia (che è una misura del disordine). L'aggregato si riorganizza in modo tale da ridurre la sua energia libera, sia rendendo minima l'energia interna, sia massimizzando l'entropia, sia con una combinazione di entrambi i processi: la materia tende verso uno stato di bassa energia interna e di entropia elevata e l'energia libera è semplicemente un indicatore del «punteggio» nella gara fra l'energia interna e l'entropia.

Si immagini ora un insieme di microaggregati mantenuti a una temperatura sufficientemente bassa da far sì che tutti si trovino in fase solida. In queste condizioni l'entropia di ciascun aggregato è bassa, perché i livelli di energia sono abbastanza distanziati e solo alcuni di essi sono occupati. L'energia libera degli aggregati assume un valore minimo perché anche l'energia interna è bassa.

Supponiamo ora di aumentare un poco la temperatura. Ciascun aggregato ha a disposizione livelli di energia più elevati e quindi un modo per aumentare la propria entropia e diminuire potenzialmente la propria energia libera. Ma se l'aumento di temperatura è modesto, l'entropia acquisita nel portarsi ai livelli di energia più elevati viene controbilanciata dal risparmio di energia interna che si ha nel rimanere a livelli bassi. Pertanto gli aggregati restano in fase solida e solo occasionalmente si avventurano nei livelli di più alta energia.

Se si continua a innalzare la temperatura, ciascun aggregato può accedere con facilità via via maggiore ai livelli di alta energia della fase liquida, livelli che sono sempre più ravvicinati sulla scala di energia crescente. Pertanto il disordine associato al singolo aggregato aumenta enormemente, dato che esso può muoversi con libertà pressoché totale fra i diversi livelli. Questa volta il grande aumento di entropia è più che sufficiente a controbilanciare l'energia interna necessaria per raggiungere i livelli della fase liquida, e quindi l'energia libera assume un valore minimo per gli aggregati che si trovano in questa fase.

In queste condizioni si ha un risultato sorprendente. Il minimo di energia libera corrispondente agli aggregati in fase solida si conserva (anche se di solito diventa meno profondo) e quindi, dal momento che vi sono due minimi di energia libera, possono coesistere aggregati in fase sia solida sia liquida.

Se si ha un ulteriore aumento di temperatura, l'equilibrio energia interna-entropia si sposta a favore degli aggregati in fase liquida e il minimo di energia libera per gli aggregati in fase solida diviene sempre meno pronunciato. Infine, quando la temperatura ha raggiunto un valore sufficientemente elevato, questo minimo scompare, lasciando solo il minimo corrispondente alla fase liquida, e l'equilibrio energia interna-entropia si sposta completamente a favore di quest'ultima fase.

Ora è possibile solo l'esistenza di aggregati in fase liquida. Dalla comparsa del liquido stabile alla scomparsa del solido stabile vi è quindi un certo intervallo di temperature nel quale possono coesistere aggregati in entrambe le fasi. Nell'intervallo di coesistenza la frazione di aggregati che si trovano in una della due fasi dipende dalla differenza dell'energia libera di queste fasi, e l'energia libera varia con la temperatura. In particolare, all'aumentare della temperatura la frazione di aggregati presenti in fase liquida segue una curva che ha la forma di una S appiattita.

Pertanto la densità dei livelli di energia e le correlazioni fra energia interna ed entropia ci hanno portato a concludere che un aggregato può avere un punto di congelamento definito, al di sotto del quale è stabile solo la fase solida, e un punto di fusione altrettanto ben determinato, al di sopra del quale è stabile solo la fase liquida. Fra i due punti vi è un intervallo finito di temperature nel quale entrambe le fasi sono stabili. In altri termini, il punto di fusione e quello di congelamento non coincidono! Abbiamo introdotto una separazione fra di essi: non è assolutamente necessario che siano alla medesima temperatura.

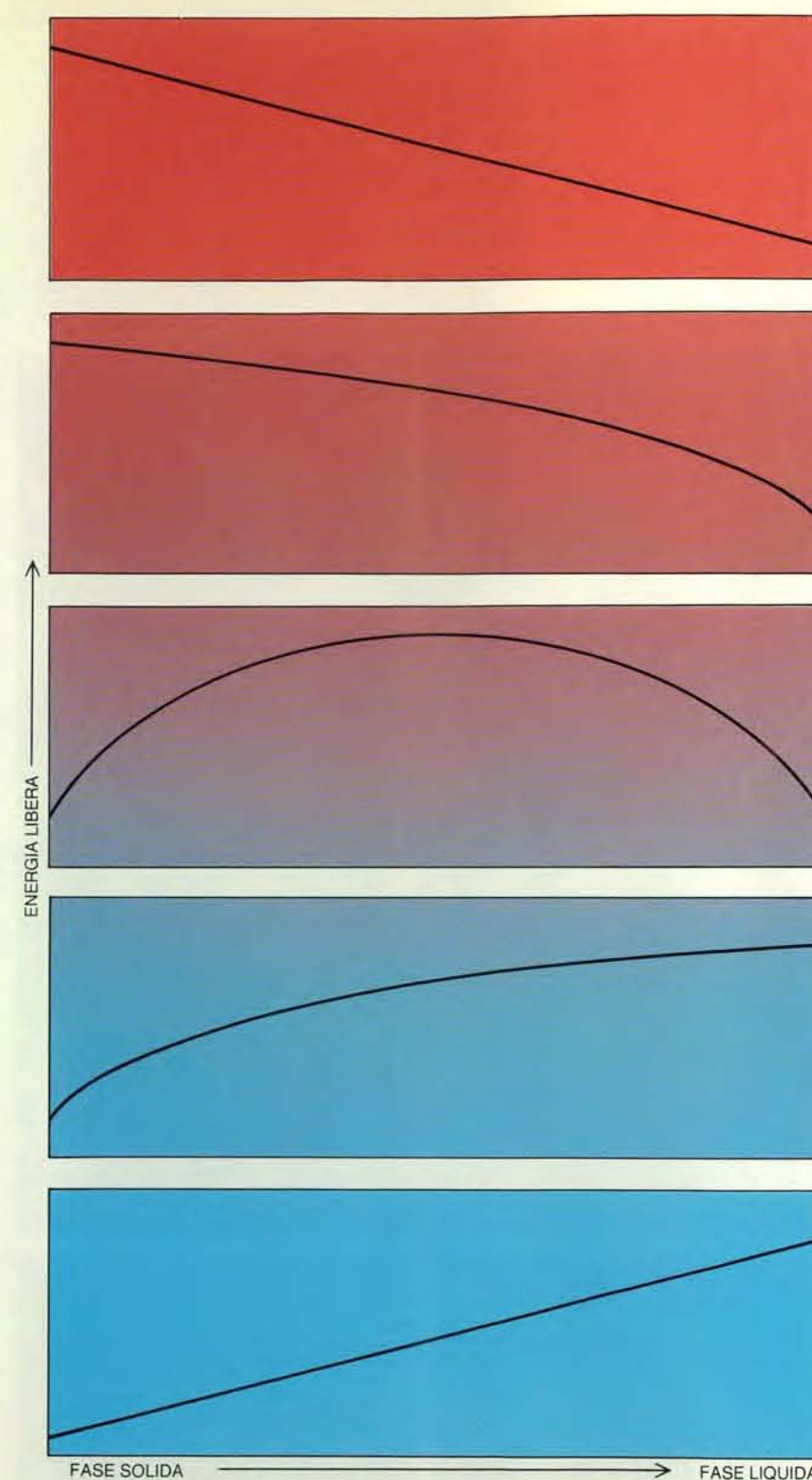
Il fatto che il punto di fusione e quello di congelamento siano distinti appare paradossale perché l'esperienza quotidiana mostra che i due valori sono uguali. Nel caso dei microaggregati, tuttavia, questa apparente contraddizione viene superata grazie alle speciali proprietà associate alle loro dimensioni intermedie. Il comportamento di un sistema piccolo - con 10, 100 o anche 10 000 atomi o molecole - è molto diverso da quello di un sistema grande, che contiene migliaia di miliardi di atomi o molecole. Per un piccolo aggregato, composto da 10-20 atomi, si prevede che la coesistenza delle fasi solida e liquida sia rilevabile in un intervallo di diversi gradi, mentre per un «superaggregato» di milioni di atomi si ritiene che l'intervallo di temperature nel quale sia possibile osservare contemporaneamente le due fasi sia inferiore a un millesimo di grado. Nel caso della materia ordinaria i limiti di questo intervallo di coesistenza sono così vicini da risultare praticamente indistinguibili, e questo giustifica l'uso intercambiabile che si fa dei termini «punto di fusione» e «punto di congelamento».

Sia gli esperimenti di laboratorio, sia le simulazioni al calcolatore forniscono dati a sostegno di questa teoria relativa alle differenze tra il punto di fusione e quello di congelamento dei microaggregati. Questi vengono sintetizzati in laboratorio, o producendo un gas degli atomi o delle molecole componenti e lasciando loro il tempo di aggregarsi, o «estraendo» semplicemente gli aggregati da un solido. Nei due casi, gli aggregati possono essere studiati come gas o come specie intrappolate in una matrice inerte.

Per distinguere gli aggregati in fase solida da quelli in fase liquida occorre essere in grado di «vedere» se essi hanno il comportamento caratteristico di uno di questi stati della materia. I solidi sono rigidi e di forma definita, tanto che, per esempio, è possibile spingerli. I liquidi, d'altra parte, sono «cedevoli» e non hanno una forma propria: non è possibile appoggiare il dito sull'acqua, ma solo immergerlo nell'acqua, dato che essa cede alla minima pressione. (Naturalmente, anche un liquido può comportarsi come un solido nei confronti di una forza esercitata all'improvviso, come sa benissimo chiunque abbia malauguratamente preso una spanciata tuffandosi in acqua.)

Un aggregato in fase solida ha una forma definita nel senso che è quasi del tutto vincolato a vibrare intorno a una singola configurazione geometrica (ossia una valle specifica). L'aggregato può vibrare milioni o anche migliaia di miliardi di volte intorno a una configurazione prima di passare a un'altra configurazione, ammesso che abbia l'energia sufficiente per farlo. (Normalmente un aggregato vibra circa 10 000 miliardi di volte al secondo.) Viceversa un aggregato in fase liquida passa facilmente da una configurazione a un'altra (ossia da una valle a un'altra). L'aggregato compie un piccolo numero di vibrazioni, forse qualche centinaio al più, prima di passare alla configurazione successiva.

Perché sia possibile osservare le fasi solida e liquida, gli aggregati devono trascorrere un tempo sufficiente in ciascuna di esse. Gli aggregati in fase liquida devono rimanervi abbastanza a lungo da mostrare le proprietà caratteristiche di un liquido, in particolare il passaggio attraverso numerose configurazioni geometriche mediante la riorganizzazione rapida degli atomi. Gli aggregati in fase solida devono presentare la rigidità e la scarsa deformabilità che sono caratteristiche di questa fase.

Attualmente i più accurati dati di laboratorio sembrano indicare che è effettivamente possibile distinguere i microaggregati in fase solida da quelli in fase liquida. Le ricerche finora compiute si sono basate sull'introduzione in ciascun aggregato di una molecola estranea, diversa da tutti gli altri atomi o molecole dell'aggregato, che funge da marcatore. Lo spettro di questa molecola (ossia la radiazione di lunghezza d'onda caratteristica emessa e assorbita dalla molecola) dipende in una certa misura dagli atomi circostanti oltre che dal mar-

La coesistenza di aggregati in fase solida e in fase liquida viene illustrata per un intervallo finito di temperature. La temperatura rappresentata dai cinque riquadri aumenta andando dal basso verso l'alto. L'asse verticale di ciascun riquadro indica l'energia libera degli aggregati, vale a dire la differenza fra la loro energia interna e il prodotto della temperatura per l'entropia. L'asse orizzontale è una misura della non rigidità degli aggregati; quelli in fase solida si trovano sulla sinistra, quelli in fase liquida sulla destra. Alla temperatura più bassa (*riquadro inferiore*) esiste un minimo di energia libera solo per gli aggregati in fase solida mentre a quella più alta (*riquadro superiore*) vi è un minimo solo per gli aggregati in fase liquida. A temperature intermedie (*i tre riquadri centrali*), tuttavia, esistono punti di minimo per gli aggregati che si trovano sia in fase solida sia in fase liquida: le due fasi coesistono. In altri termini, il punto di fusione e quello di congelamento non coincidono.

catore stesso. Di conseguenza lo spettro di un marcatore introdotto in un aggregato in fase solida dovrebbe essere diverso da quello di un marcatore introdotto in un aggregato in fase liquida.

L'impiego di un laser per eccitare la molecola marcatore a un livello di energia noto, è un metodo estremamente efficace per identificare un aggregato. (La lunghezza d'onda della radiazione laser necessaria per indurre questa eccitazione dipende strettamente dal tipo di marcatore e dagli atomi che lo circondano nell'aggregato.) Poi un secondo laser viene sintonizzato a una lunghezza d'onda tale da provocare la ionizzazione della molecola marcatore già eccitata: ciò significa che il secondo laser strappa alla molecola uno dei suoi elettroni esterni. L'elettrone liberato sfugge, lasciando una carica positiva sul marcatore e sull'aggregato in cui esso è inserito. La carica risultante offre la possibilità di applicare una forza che acceleri l'aggregato (mediante un campo elettrico) e quindi lo estragga da un miscuglio. L'aggregato viene quindi identificato in base alla sua massa, ossia al numero di atomi o molecole che contiene (si veda l'articolo *Il rilevamento con il laser di atomi e molecole* di Vladilen S. Letokhov in «Le Scienze» n. 243, novembre 1988).

Jürgen Bösiger e Samuel Leutwyler dell'Università di Berna e Mee Hahn e Robert L. Whetten dell'Università della California a Los Angeles hanno utilizzato questo metodo per studiare aggregati di atomi di argo. I ricercatori svizzeri hanno impiegato come marcatore una molecola di carbazolo, dalla forma piatta e allargata, mentre il gruppo statunitense si è servito del benzene, che è più piccolo e tondeggiante. Le molecole marcatore introdotte negli aggregati in fase solida presentano uno spettro netto e distinto, mentre quelle inserite negli aggregati in fase liquida hanno uno spettro allargato. (Nonostante le sottigliezze riguardanti l'interpretazione dei dati sperimentali, le argomentazioni che evidenziano le differenti caratteristiche dei due tipi di spettri appaiono convincenti.)

Marcatori inseriti negli aggregati permettono di distinguere la fase liquida da quella solida. In alto, un atomo di sodio che funge da marcatore (*in rosso*) è stato inserito in un aggregato di argo con 54 atomi. L'immagine è stata realizzata da Hai-Ping Cheng dell'Università di Chicago. Il marcatore emette radiazione di lunghezza d'onda diversa a seconda che l'aggregato sia in fase solida (*in basso a sinistra*) o in fase liquida (*in basso a destra*).

Le due serie di esperimenti sembrano indicare che a una temperatura compresa tra 20 e 30 kelvin (da −253 a −243 gradi Celsius) i microaggregati piccoli hanno un comportamento simile a quello di un liquido (con spettri allargati) mentre quelli grandi appaiono più simili a un solido (e hanno spettri netti). Gli spettri dei microaggregati di dimensione intermedia indicano la coesistenza di aggregati in fase solida e in fase liquida.

I dati ottenuti da questi laboratori rappresentano un importante passo avanti nel programma sperimentale teso alla conoscenza degli aggregati. La coesistenza di aggregati in fase solida e liquida in un intervallo finito di dimensioni, ma a temperatura costante, indica senz'altro che il punto di fusione e quello di congelamento devono essere differenti, anche se non si tratta di una dimostrazione completa. Idealmente occorrerebbe mantenere costante la dimensione degli aggregati e far variare la temperatura, ma un tale esperimento richiederebbe una precisione molto difficile da raggiungere. Per dimostrare inconfutabilmente questa conclusione saranno necessarie ulteriori ricerche sperimentali.

Un altro metodo per verificare la teoria è quello di realizzare simulazioni al calcolatore degli aggregati. Naturalmente una simulazione non può sostituire un esperimento reale, ma, nel momento in cui vi siano modi oggettivi per stabilirne la validità, può rivelare fenomeni nuovi e consentire «esperimenti» di gran lunga meno costosi di quelli effettuati in laboratorio. I modelli al calcolatore possono anche essere di aiuto

Le simulazioni al calcolatore dell'evoluzione della temperatura media degli aggregati di argo con 13 atomi a diversi valori di energia costante mostrano la coesistenza di fasi liquide e solide. A bassa energia (*in alto*) gli aggregati si comportano come solidi, ad alta energia (*al centro*) come liquidi e a energia intermedia (*in basso*) si comportano sia come solidi sia come liquidi. La temperatura dell'aggregato è direttamente proporzionale alla sua energia cinetica. Ciascun puntino bianco rappresenta 5×10^{-12} secondi.

nell'ideare esperimenti di laboratorio.

I casi in cui risulta particolarmente utile sfruttare la potenza del calcolatore per simulare un sistema sono quelli dove esiste un valido modello teorico, di applicazione però troppo complessa per essere eseguita a mano: un esempio è proprio il comportamento di un aggregato di atomi. Le forze che agiscono fra due atomi sono ben note nel caso di molte sostanze, ma è impossibile risolvere analiticamente le equazioni del moto anche solo per tre atomi interagenti (per non parlare delle situazioni in cui gli atomi sono più di tre). Le equazioni relative al moto di diversi atomi (anche di un numero molto grande) possono tuttavia essere risolte numericamente con l'aiuto di un calcolatore. In questo modo nella simulazione si può seguire il comportamento di ciascun «atomo».

Esaminiamo ora brevemente le simulazioni di aggregati di atomi di argo eseguite da Thomas L. Beck, Hai-Ping Cheng, Heidi L. Davis, Julius Jellinek, David J. Wales e da me all'Università di Chicago. Di tutte le informazioni fornite da queste simulazioni, la più facile da comprendere è la storia di un singolo aggregato che si evolve nel tempo. Una grandezza molto utile che può essere ricavata facilmente è la temperatura media di un aggregato la cui energia totale sia costante o, viceversa, l'energia media di un aggregato la cui temperatura sia costante. (La temperatura media è pari, a meno di un fattore numerico di conversione, all'energia cinetica media.)

Il significato della parola «medio» in questi esperimenti è fondamentale. Calcolare la media della temperatura una sola volta in tutta la storia di un aggregato può essere altrettanto fuorviante che calcolarla per un tempo così breve che ciascun atomo dell'aggregato non abbia quasi la possibilità di interagire con gli altri atomi. In genere calcoliamo la media su cinque milionesimi di milionesimo di secondo, un tempo sufficiente perché gli atomi di argo compiano solo alcune vibrazioni.

Abbiamo scoperto che un aggregato di 13 atomi di argo mantenuto in una situazione di bassa energia presenta minuscole fluttuazioni di temperatura intorno alla media e mostra un comportamento simile a quello di un solido. Un aggregato di energia molto più elevata ha una temperatura media significativamente più alta e fluttuazioni analogamente più ampie intorno a questa temperatura. Diversi aspetti del suo comportamento indicano che l'aggregato è in fase liquida. Se si aumenta a sufficienza l'energia dell'aggregato, la simulazione mostra atomi che «evaporano» da esso, proprio come ci si aspetterebbe.

Fra i due estremi vi è un intervallo di energie in cui alcuni aggregati, di energia costante, presentano un comportamento davvero notevole: trascorrono lunghi periodi di tempo in fase solida, poi, più o meno casualmente, passano in fase liquida rimanendovi a lungo, quindi tornano in fase solida e così via. Questa coesistenza è una caratteristica anche di aggregati dell'argo composti da sette, 15 o 19 atomi.

È però opportuno notare che non tutti gli aggregati mostrano chiaramente la coesistenza di due fasi: un esempio di questo tipo sono gli aggregati dell'argo costituiti da sei, otto o 17 atomi. Gli aggregati di sei atomi di rame presentano fasi solide sia rigide sia «morbide». Alcuni aggregati di cloruro di sodio, secondo simulazioni condotte da Daphna Scharf e Joshua Jortner dell'Università di Tel Aviv e da Uzi Landman del Georgia Institute of Technology, possono avere simultaneamente regioni in fase solida e in fase liquida su lati diversi dello stesso aggregato. In simulazioni effettuate da Cheng, le superfici di aggregati di argo composti da 52 o 55 atomi fondono a temperature molto inferiori alla temperatura di fusione degli aggregati stessi; questa osservazione dimostra che negli aggregati le caratteristiche superficiali hanno un'importanza molto maggiore che nella materia convenzionale.

In definitiva, i microaggregati possono mostrare una grande varietà di comportamenti via via che si aumenta la loro energia e che la loro struttura diventa meno rigida e più deformabile. Certamente questa complessità di fenomeni si rivelerà uno strumento utile per studiare sempre più a fondo i processi di transizione di fase.

BIBLIOGRAFIA

HILL T. L., *The Thermodynamics of Small Systems*, parte 1 e 2, Benjamin Co., Inc., 1963.

Large Finite Systems in *Proceedings of the Jerusalem Symposia on Quantum Chemistry and Biochemistry*, 20, a cura di Joshua Jortner, Alberte Pullman e Bernard Pullman, Reidel Publishing Co., 1987.

SUGANO S., NISHINA Y. e OHNISHI S. (a cura), *Microclusters*, Springer-Verlag, 1987.

JENA P., RAO B. K. e KHANNA S. N. (a cura), *Physics and Chemistry of Small Clusters*, Plenum Press, 1987.

BENEDEK G. e PACCHIONI MARTIN (a cura), *Elemental and Molecular Clusters*, Springer-Verlag, 1988.

Evolution of Size Effects in Chemical Dynamics, parte II, in *Advances in Chemical Physics*, 70, a cura di Ilya Prigogine e Stuart A. Rice, Wiley-Interscience, 1988.

The Chemical Physics of Atomic and Molecular Clusters in *Proceedings of the CVII Course of the International School of Physics «Enrico Fermi»*, Varenna, 289, a cura di Giacinto Scoles e S. Stringari, Società Italiana di Fisica, in stampa.

La dispersione dei semi da parte delle formiche

Migliaia di specie vegetali distribuite in tutto il mondo dipendono da questi insetti per la disseminazione: li attraggono con esche alimentari o altre strategie e li inducono a diffondere i semi senza danneggiarli

di Steven N. Handel e Andrew J. Beattie

Spesso le piante devono ricorrere alla collaborazione degli animali per disperdere i propri semi su un vasto areale. Alcune specie producono semi che si attaccano temporaneamente al mantello dei mammiferi o al piumaggio degli uccelli; gli alberi da frutto e altre piante hanno invece semi che vengono ingeriti dagli animali frugivori, i quali li rigurgitano o li eliminano con gli escrementi, facendo così in modo che questi semi possano germinare a distanza. Non sono però solo i vertebrati a disperdere i semi delle piante: anche le formiche svolgono un ruolo importante.

I biologi stanno appena cominciando a riconoscere i meccanismi specializzati che fanno della formica uno dei principali artefici della diffusione delle piante in tutto il mondo. E, in effetti, le piante che si avvalgono di questo imenottero per la dispersione dei semi si trovano in habitat diversi su tutti i continenti, a eccezione dell'Antartide. Oggi sono note più di 3000 specie di angiosperme, appartenenti a oltre 60 famiglie, i cui semi vengono dispersi in questo modo e la lista è ben lungi dall'essere completa.

Tra molte piante e le formiche che ne trasportano i semi esiste un vero e proprio mutualismo, cioè un'interazione vantaggiosa per ambo le parti. Questo mutualismo è sorto indipendentemente in un numero così elevato di gruppi di piante da far pensare all'esistenza di forti pressioni, frequentemente ricorrenti, che lo favoriscano. In questo articolo concentreremo la nostra attenzione sull'evoluzione di questo mutualismo e sui vantaggi ecologici che esso ha prodotto.

Le formiche disperdono i semi mediante due meccanismi generali. Il primo dipende dalla scarsa efficienza con la quale le formiche foraggiatrici, che raccolgono grandi quantità di semi, trasportano questi ai nidi per poi nutrirsene. Alcuni semi cadono e vanno perduti lungo il tragitto, altri possono rimanere nascosti nel terreno, in modo da sfuggire alle formiche operaie. Germinando, questi semi possono dare origine a plantule che colonizzano altre località.

Poiché il numero dei semi mangiati dalle formiche supera quello dei semi perduti o ignorati, questo meccanismo di dispersione sembrerebbe più vantaggioso per le formiche che non per le piante, a cui viene in definitiva sottratta gran parte dei semi. Il meccanismo di dispersione da parte delle formiche foraggiatrici viene allora descritto più come un esempio di predazione che non di mutualismo. Il suo impatto è limitato in larga misura agli habitat aridi.

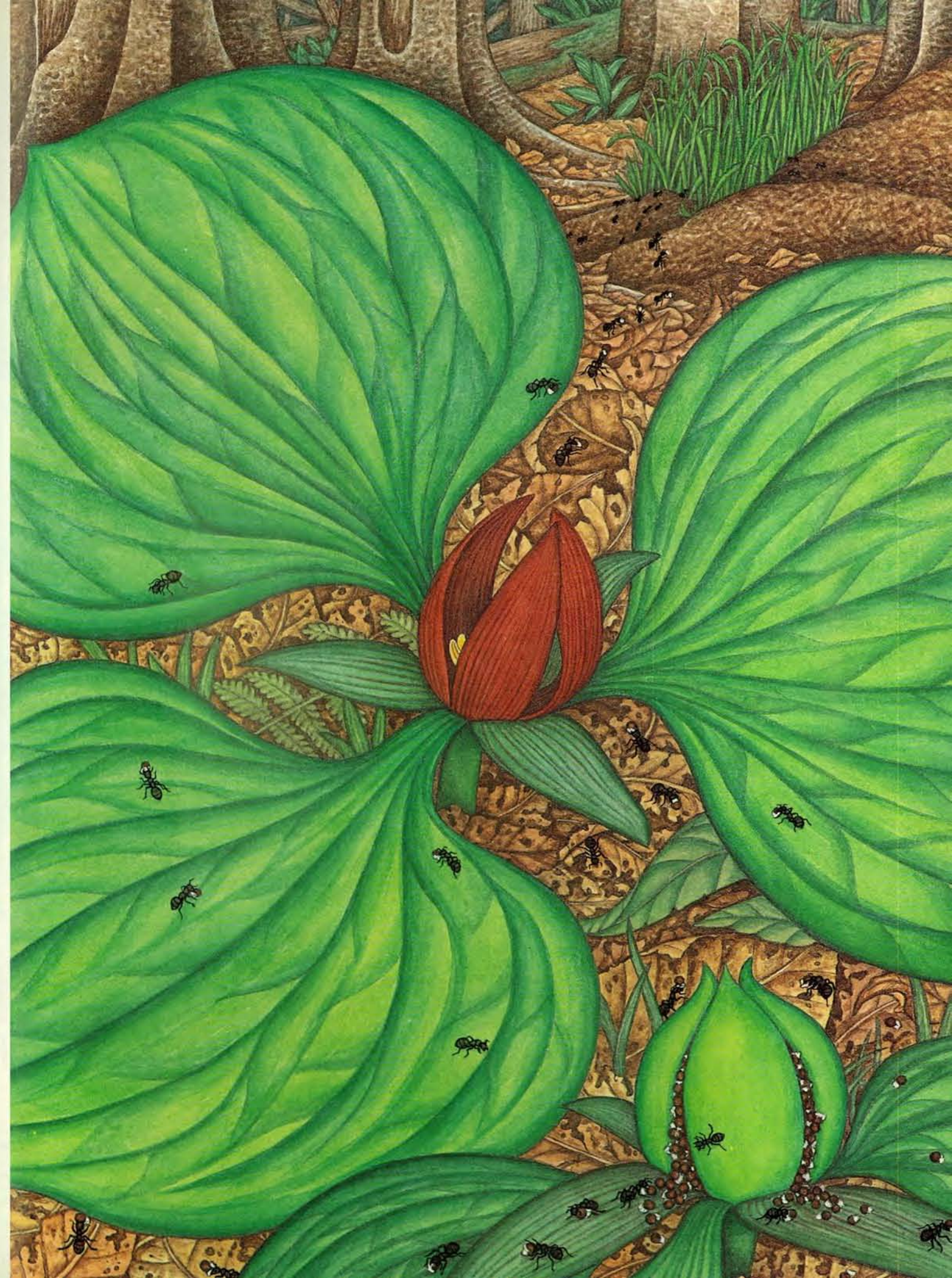
Il secondo meccanismo di dispersione, che ci interessa maggiormente, è diverso e ha un significato di gran lunga più ampio. Riguarda le piante che producono un elaiosoma (o eleosoma), cioè un corpo ricco di grassi attaccato al seme o vicino a esso. Gli elaiosomi attraggono le formiche, che li trasportano, assieme ai semi, nei propri nidi; qui i membri della colonia mangiano gli elaiosomi e scartano i semi senza danneggiarli.

Questo meccanismo non sacrifica i semi alla predazione delle formiche, ma è un esempio di reale mutualismo, che porta beneficio sia alle formiche sia alle piante. Lo si definisce «mirmecocoria», un termine un po' impropriamente mutuato dal greco che significa appunto «dispersione a opera delle formiche».

Gli elaiosomi sono comparsi a più riprese, con la funzione di esche per le formiche, durante l'evoluzione di diverse famiglie di piante. Essi si trovano comunemente nelle piante dei boschi umidi dell'America nordorientale e dell'Europa, nelle comunità vegetali della boscaglia di ambiente arido dell'Australia orientale e nelle comunità vegetali dell'Africa meridionale.

Spesso solo alcune specie in ogni famiglia si affidano alle formiche per la dispersione dei semi. Per fare un esempio, nel genere *Carex*, ricco di specie, solo alcune portano elaiosomi di cui sia stata dimostrata la funzione di esca per le formiche. In molte altre specie dello stesso genere la dispersione avviene per mezzo dell'acqua o di vertebrati. Analogamente nel genere *Trillium*, comprendente gigliacee caratterizzate da grossi fiori, che crescono nei boschi, alcune specie hanno semi con elaiosomi, che vengono dispersi dalle formiche, mentre altre hanno frutti carnosi contenenti semi che vengono dispersi dai vertebrati. Come illustrano questi esempi tratti da gruppi filogeneticamente molto diversi, la mirmecocoria può comparire indipendentemente nell'ambito di vari generi.

Questo meccanismo di dispersione dei semi fu studiato per la prima volta in modo esauriente da Johan Rutger Sernander, botanico dell'Università di Uppsala in Svezia, il quale nel 1906 pubblicò una rassegna sulle piante mirmecocore europee. Per la ragguardevole impostazione sperimentale, quantitativa, della ricerca, il libro di Sernander servì a stabilire quale fosse l'importanza

Questa gigliacea selvatica, *Trillium petiolatum*, si è evoluta con una morfologia molto caratteristica, che stimola le formiche a disperderne i semi. Il fiore spunta insolitamente vicino al suolo, il che permette alle formiche di raggiungere più facilmente i semi maturi. Come ulteriore fattore stimolante, i semi sono dotati di corpi ricchi di grassi, gli elaiosomi, che esercitano sulle formiche una forte attrazione. Le formiche trasportano nei loro nidi i semi con gli elaiosomi dei quali si nutrono scartando la parte germinale che finirà per svilupparsi.

Dicentra cucullaria (*a sinistra*) **è una pianta erbacea che cresce nelle regioni boscose e produce semi con elaiosomi che maturano ed escono dall'involucro (*al centro*) verso la fine della primavera. A maggiore ingrandimento (*a destra*), vengono mostrati quattro semi muniti di elaiosomi insieme ad altri quattro semi prelevati da un cumulo di rifiuti in un formicaio. Questi ultimi, pur essendo privi degli elaiosomi che sono stati ingeriti dalle formiche, trovano un ambiente molto adatto alla germinazione nel fertile suolo del nido.**

della mirmecocoria per buona parte della vegetazione europea. Egli aveva eseguito una lunga serie di esperimenti sul campo, prendendo in considerazione molte specie vegetali e dimostrando così che le formiche, quando possono scegliere tra una varietà di semi, preferiscono diffondere quelli provvisti di elaiosomi.

Le piante mirmecocore furono studiate in un primo tempo in Europa, ma ben presto i botanici cominciarono a cercare questo meccanismo di dispersione dei semi nella vegetazione degli altri continenti. Poco per volta, alla lista delle mirmecocore note si aggiunsero specie del Nuovo Mondo. In Europa e nell'America Settentrionale, le mirmecocore sono generalmente piante erbacee che crescono nelle foreste umide, decidue. (Sernander fu il primo a notare questa correlazione.) Nell'America Latina, i semi delle piante erbacee, delle epifite e delle liane della foresta pluviale tropicale vengono spesso dispersi a opera delle formiche.

Le piante mirmecocore sono particolarmente abbondanti anche in Australia e nell'Africa meridionale, dove sono rappresentate da arbusti a foglie coriacee che crescono su un suolo arido e povero di sostanze nutritive. Nel 1975 Rolf Y. Berg dell'Università di Oslo ha pubblicato i risultati di ricerche secondo cui i semi di circa 1500 specie di piante australiane, appartenenti a 87 generi diversi, vengono dispersi dalle formiche. La vegetazione chiamata *fynbos*, nell'Africa meridionale, contiene anch'essa più di 1000 specie di questo tipo. Le ricerche oggi in corso nelle regioni tropicali aggiungeranno indubbiamente molti altri nomi alla lista.

La diversità tassonomica delle piante provviste di elaiosomi trova una corrispondenza nell'ampia varietà di tessuti vegetali che l'evoluzione ha modificato in strutture adatte ad attrarre le formiche. In alcune specie, per esempio in *Dicentra cucullaria*, una sezione del tessuto che costituisce il rivestimento del seme si è espansa e modificata a formare un elaiosoma. In altre, tra le quali le ranuncolacee a fioritura primaverile del genere *Hepatica*, che crescono nell'America nordorientale, l'elaiosoma deriva da una porzione della parete dell'ovario che circonda il seme. In certe specie di *Carex* gli elaiosomi hanno origine invece dal tessuto bratteale che circonda l'ovario. Si conoscono numerose altre strutture che producono elaiosomi nelle angiosperme.

Nel complesso, le origini diverse degli elaiosomi costituiscono un buon esempio di evoluzione convergente, che dimostra come strutture con forme e funzioni differenti possano essere rimodellate dalla selezione naturale per svolgere un medesimo ruolo ecologico. Nel caso degli elaiosomi, tessuti vegetali aventi in origine una funzione protettiva contro gli insetti erbivori e altre minacce sono stati modificati biochimicamente e strutturalmente in esche commestibili per le formiche.

Gli elaiosomi sono composti da cellule profondamente modificate che contengono grossi vacuoli riempiti da un ricco miscuglio di sostanze nutritive. Andreas Bresinsky, dell'Università di Monaco di Baviera, dopo aver esaminato una vasta selezione di piante mirmecocore, ha riferito che gli elaiosomi contengono molti grassi, acidi grassi e altre comuni sostanze nutritive indispensabili per gli animali. Queste sostanze potrebbero dunque essere consumate come cibo dalle formiche.

La maggior parte delle formiche è onnivora e mangia insetti e parti di piante e di animali che trova sul suolo. È possibile che gli elaiosomi e i semi a essi uniti «imitino» chimicamente i tessuti animali e in questo modo traggano in inganno le formiche.

Gli elaiosomi potrebbero contenere anche sostanze stimolanti che innescano nelle formiche il comportamento di foraggiamento. Diane L. Marshall dell'Università del New Mexico, assieme ai suoi collaboratori, ha identificato negli elaiosomi della viola mammola (*Viola odorata*) una sostanza specifica, l'1,2-dioleina, che è un composto lipidico polare. Un lipide analogo è stato scoperto negli elaiosomi di due arbusti australiani, *Acacia myrtifolia* e *Tetratheca stenocarpa*.

L'importanza di questi composti per le formiche non è ancora chiara, ma la loro presenza in piante mirmecocore che crescono in regioni del globo così lontane l'una dall'altra fa pensare che sia in atto una convergenza evolutiva. Fatto ancora più interessante, questa somiglianza conduce all'ipotesi che l'elaiosoma provochi altri comportamenti del resto già innati nelle formiche oltre a quello di raccolta del cibo. Per esempio, si sa che l'acido oleico stimola la necroforesi, ossia il trasporto di corpi morti, in alcune formiche; si pensa, dunque, che anche gli elaiosomi contenenti questa sostanza vengano raccolti per la stessa ragione.

Oltre agli elaiosomi che fungono da esche alimentari, le piante mirmecocore presentano talvolta altri adattamenti morfologici che facilitano il piazzamento dei semi presso le piste percorse dalle formiche. In alcune specie sia lo stelo sia il sostegno del frutto sono esili e deboli e si piegano quando i semi sono maturi. Questo fa sì che i semi vengano a trovarsi in prossimità del suolo, o addirittura a contatto con esso, proprio dove le formiche vanno alla ricerca di cibo.

In altre specie vegetali la morfologia tipica subisce alterazioni più profonde. Per esempio, in *Carex umbellata*, una specie di pianta palustre da noi studiata, lo stelo fiorale è eccezionalmente corto e i semi, con i tessuti circostanti, maturano in prossimità del suolo. Di conseguenza, invece di svilupparsi su lunghi steli che si piegano verso terra una volta raggiunta la maturità, i semi rimangono per tutto il periodo di sviluppo stagionale a portata delle formiche.

Una alterazione ancora più marcata si osserva nella morfologia di *Trillium petiolatum*, una gigliacea selvatica dell'America nordoccidentale. La maggior parte delle piante del genere *Trillium* ha un fiore e tre foglie che si sviluppano alla sommità di un lungo stelo, spesso alto 30 centimetri. In *T. petiolatum*, invece, il fiore grosso e vistoso si forma in prossimità del suolo, dove i semi con elaiosomi maturano e finiscono alla portata delle formiche.

Inoltre nel genere *Trillium* le foglie hanno per lo più una disposizione spiralata al di sotto del fiore, ma in *T. petiolatum* questa disposizione farebbe arrivare le foglie a livello del pavimento della foresta. Pertanto le lamine fogliari, pur originando in posizione normale sotto il fiore, si sviluppano all'estremità di lunghi piccioli o steli che le rialzano al di sopra del fiore, in una posizione più favorevole per la fotosintesi. In breve, la struttura usuale del genere *Trillium* è qui pressoché capovolta. Una spiegazione evoluzionistica ragionevole della conformazione di *T. petiolatum* deve tener conto dei notevoli vantaggi che sono legati alla dispersione dei semi da parte delle formiche.

Per facilitare la disseminazione, le piante mirmecocore potrebbero anche avere alterato i tempi di maturazione dei propri semi. Per esempio, nei climi temperati la maggior parte di esse presenta semi ed elaiosomi che maturano all'inizio della primavera. In questo periodo gli insetti morti, che rappresentano spesso la fonte di cibo primaria delle formiche, non sono così abbondanti come in estate, dopo che è avvenuta la riproduzione. Pertanto le piante i cui elaiosomi maturano in primavera possono attrarre più facilmente l'attenzione delle formiche che ricercano il cibo e i semi potrebbero venire asportati con una frequenza maggiore di quelli prodotti in estate o in autunno.

La selezione naturale in favore di una maturazione precoce dei semi e degli elaiosomi potrebbe spiegare così la predominanza di piante mirmecocore a fioritura primaverile. Certamente possono esservi altri fattori a favorire l'elevata attività metabolica delle piante erbacee boschive in primavera: per esempio la maggiore disponibilità di luce a livello del pavimento forestale prima che si sviluppino le fronde della volta. Lo sfruttamento della ricerca di cibo da parte delle formiche potrebbe essere soltanto un'ulteriore pressione selettiva che ha favorito l'evoluzione delle piante mirmecocore i cui semi maturano all'inizio della primavera.

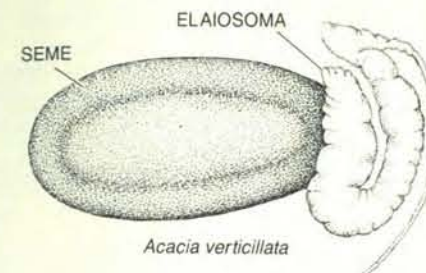
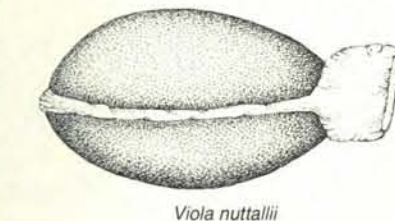
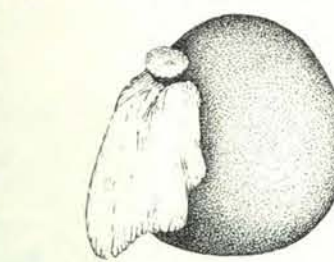
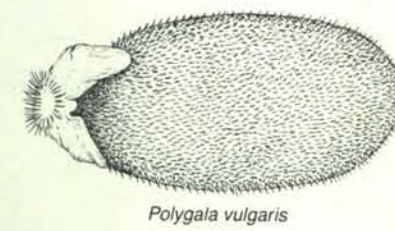
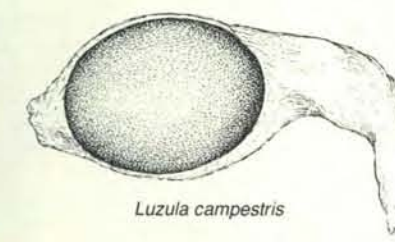
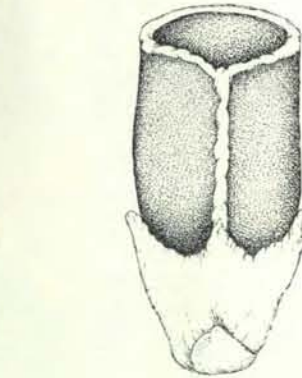
Le formiche che raccolgono i semi delle piante mirmecocore sono molte e di specie assai diverse. Parecchie di esse sono chiaramente adattate a un tipo di alimentazione che è in primo luogo carnivora. Per esempio, Carol C. Horvitz dell'Università di Miami ha dimostrato che nel Messico meridionale le formiche che trasportano i semi di *Calathea* appartengono ai generi *Odontomachus* e *Pachycondyla*, caratterizzati da robusti pungiglioni e da grosse mandibole per sopraffare la preda. Le formiche raccolgono i semi con grande avidità e li portano nei loro nidi dove staccano gli elaiosomi per darli in pasto alle larve. Forse alcune sostanze chimiche presenti negli elaiosomi stimolano le formiche nello stesso modo in cui le stimolerebbe una preda.

Molti altri generi di formiche usano disperdere semi: *Formica*, *Myrmica* e *Aphaenogaster* sono comuni nelle foreste temperate europee e nordamericane; specie di *Rhytidoponera*, *Pheidole* e *Iridomyrmex* sono numerose nelle brughiere dell'Australia sudorientale. Perfino *Messor*, *Pogonomyrmex* e *Veromessor*, che normalmente raccolgono semi e se ne nutrono, in certe condizioni sono state identificate come agenti di disseminazione.

Da un punto di vista strategico, è importante che la mirmecocoria interessi quante più specie di formiche possibile. In qualunque ambiente si trovano generalmente molte specie diverse e una pianta che si sia evoluta in modo da attrarre soltanto una si troverebbe probabilmente in svantaggio. Di fatto, delle migliaia di specie di piante mirmecocore note, non una è risultata dipendente da una singola specie di formiche.

Allo stesso modo non vi è alcuna prova che una qualsiasi specie di formiche si sia evoluta in modo da ricavare vantaggio da una particolare pianta mirmecocora. Questa mancanza di specializzazione è in netto contrasto con le numerose relazioni specie-specifiche tra insetti e piante che sono state osservate ai tropici, relazioni spesso assai importanti per l'impollinazione.

Pertanto, il fenomeno della mirmecocoria si può descrivere più propriamente come il risultato di un'evoluzione delle piante, e non come quello di una coevoluzione di una pianta e di un insetto. Per le formiche un elaiosoma potrebbe esse-

ALCUNI SEMI CON ELAIOSOMI

Acacia verticillata

Viola nuttallii

Corydalis aurea

Polygala vulgaris

Luzula campestris

Lamium album

re nient'altro che una fonte di sostanze nutritive da trasportare nel nido.

Perché proprio le formiche sono state prescelte dalla selezione naturale per effettuare la disseminazione? Negli habitat in cui crescono le piante mirmecocore sono certamente comuni molti altri gruppi di insetti. Ma perché la dispersione dei semi si svolga in modo efficace è necessario che un insetto li trasferisca a una distanza significativa senza danneggiarli. Solo gli insetti sociali, che portano il cibo al nido invece di mangiarselo sul posto, soddisfano entrambi i criteri. Le formiche operaie ripuliscono in genere l'area attorno al nido e trasportano il cibo nel nido stesso per nutrire le larve. L'evoluzione del comportamento sociale ha pertanto preadattato, o predisposto, le formiche a diventare agenti ideali per la dispersione dei semi.

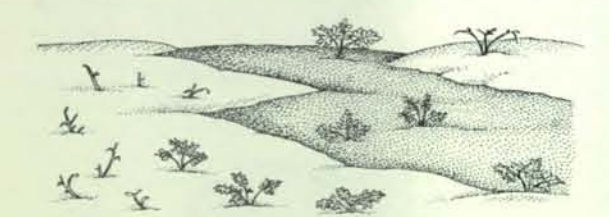
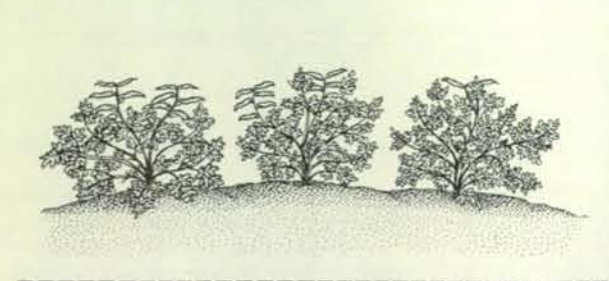
Le formiche hanno anche caratteristiche comportamentali che le rendono particolarmente adatte a svolgere il ruolo di agenti di dispersione dei semi. Nella maggior parte degli habitat esse sono gli insetti più abbondanti. Svolgono la loro attività in superficie mostrandosi notevolmente efficienti nella raccolta del cibo durante la lunga stagione di crescita delle piante. Quando trovano nuove risorse alimentari, reclutano operaie per raccogliere quanto più cibo possibile. Arrivano persino al punto di trasferire l'intero nido in aree che sembrano particolarmente ricche di fonti alimentari. Tutti questi comportamenti sono evidentemente vantaggiosi per le piante mirmecocore.

Dal momento che la mirmecocoria è presente in habitat molto diversi un po' in tutto il mondo, gli ecologi si sono chiesti se sia in qualche modo possibile trarre conclusioni generali sui vantaggi evolutivi di questo fenomeno per le piante. Solo di recente una serie di esperimenti sul campo e in laboratorio ha messo in luce come la funzione adescatrice esercitata dai semi nei riguardi delle formiche foraggiatrici possa accrescere la sopravvivenza e la fecondità di una specie vegetale.

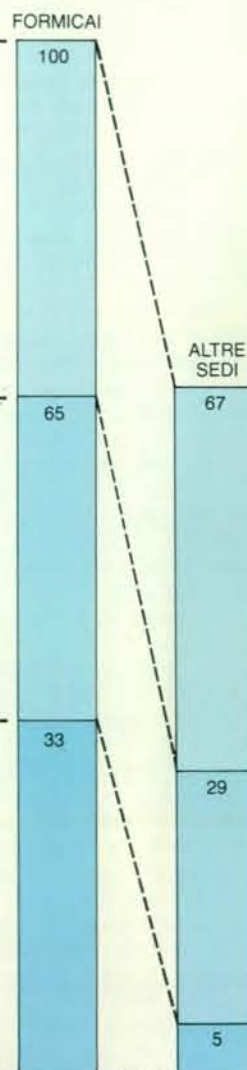
Un vantaggio fondamentale legato alla dispersione dei semi da parte delle formiche è l'espansione dell'area di distribuzione della pianta che ha prodotto i semi. Sono state spesso osservate formiche che trasportavano i semi per un metro o due, ma sono stati anche riferiti spostamenti di 70 metri. Pertanto, le formiche rendono possibile la colonizzazione di nuove aree da parte delle specie vegetali. Disperdendo la propria popolazione, una specie vegetale riduce il rischio di estinguersi a seguito di cambiamenti locali nell'habitat. Qualunque specie di formica, indipendentemente dalle modalità con cui costruisce il nido, può dare questo vantaggio.

Le formiche possono anche fare aumentare le probabilità di sopravvivenza di un seme trasportandolo lontano dalla pianta che lo ha generato: la stessa pianta, infatti, facendogli ombra potrebbe inibirne lo sviluppo. In un esperimento eseguito da uno di noi (Handel), i semi di *Carex pedunculata* lasciati crescere all'ombra della pianta madre hanno dato origine a plantule con tre sole foglie. Nello stesso periodo, le plantule sviluppatesi da semi che erano stati trasportati lontano dalla pianta madre hanno prodotto una media di 89 foglie. Inoltre questi semi erano di gran lunga più fecondi: i soli che abbiano dato origine a piante fiorite l'estate successiva.

VANTAGGI DELLO SVILUPPO SU FORMICAI

Due gruppi di 2550 semi di *Corydalis aurea* sono stati piantati in formicai e in altre sedi. L'istogramma mostra il successo relativo delle piante di ciascun gruppo in stadi differenti. Un numero di gran lunga superiore di semi è stato prodotto dalle piante cresciute sui formicai.

NUMERO TOTALE DI SEMI PRODOTTI

FORMICAI: 20 448

ALTRE SEDI: 10 718

Le formiche operaie della specie *Aphaenogaster rudis* trasportano le larve in prossimità di semi di *Sanguinaria canadensis*, in modo che queste possano nutrirsi degli elaiosomi. Non si sa ancora esattamente quali vantaggi le formiche traggano da questo alimento.

Il trasporto dei semi da parte delle formiche, oltre a contribuire a minimizzare la competizione tra pianta madre e plantule, può ridurre anche la competizione tra differenti specie vegetali. Per esempio, uno di noi (Handel) ha effettuato esperimenti con tre specie di *Carex* che crescevano nello stesso habitat, una sola delle quali era mirmecocora. Questa non cresceva bene in presenza delle altre, ma prosperava quando era isolata.

Dato che le formiche nella zona dell'esperimento erano attratte soltanto da semi provvisti di elaiosomi, finivano per provocare un isolamento naturale di questi semi nei loro nidi. La pianta mirmecocora riusciva dunque a monopolizzare certi siti dell'habitat in cui i nidi di formiche erano comuni, per esempio il legno marcescente. Essa non doveva competere con le plantule delle altre specie di *Carex* per lo spazio, la luce, le sostanze nutritive e altre risorse fondamentali. La mirmecocoria poteva essere una strategia efficace anche in presenza di molti altri generi di vegetali le cui plantule competono per i siti di crescita.

Ancor più della competizione, la predazione è una causa importante della riduzione numerica di semi e plantule. Molti animali, tra cui uccelli e piccoli roditori, mangiano soprattutto semi. E chiunque pratica il giardinaggio sa che le lumache e le limacce sono una minaccia mortale per le giovani plantule.

In diverse parti del mondo la ricerca si è concentrata sulla possibilità che i semi portati nei nidi delle formiche sfuggano almeno ad alcuni predatori. Studi effettuati nelle foreste della Virginia occidentale e nei prati di montagna del Colorado hanno rivelato che i semi posti in piccole zone da cui le formiche erano escluse venivano invariabilmente mangiati nell'arco di 24 ore. Per contro, in presenza di formiche, i semi con gli elaiosomi venivano portati rapidamente sottoterra. Nel caso di *Viola nuttallii*, che cresce nel Colorado, Christine Turnbull della Macquarie University ha dimostrato che la liberazione stagionale e giornaliera dei semi coincide con il livello massimo di attività delle formiche.

In Australia si segnala un esempio particolarmente interessante di strategia antipredatoria. Nelle regioni boscose e nelle savane, dove la vegetazione è per lo più costituita da arbusti sclerofilli (a foglie coriacee), le specie mirmecocore sono numerose. E pure lo sono i predatori ma, per ironia della sorte, le principali specie predatrici sono formiche. Un recente lavoro di Lesley Hughes, anch'egli della Macquarie University, suggerisce che i semi liberati in questo ambiente attendano i vincitori di una «gara» tra specie di formiche benefiche, che disperdono i semi, e specie dannose, che se li mangiano. Il destino dei semi dipende da quale tipo di formiche li trovi per primo. La presenza di elaiosomi sui semi accresce la probabilità che una formica disseminatrice raggiunga i semi stessi prima di una formica predatrice.

Un'altra minaccia è costituita dagli incendi, particolarmente nella vegetazione australiana e sudafricana dove predominano gli arbusti. In queste piante, però, sono comuni gli adattamenti per sopravvivere al fuoco. Inoltre molte specie, alcune delle quali sono mirmecocore, non sono semplicemente resistenti al fuoco, ma dipendono addirittura dagli incendi per potersi riprodurre.

Ricerche effettuate da diversi autori australiani indicano che la rimozione dei semi da parte delle formiche e il loro trasporto nei nidi proteggono i semi stessi dal calore distruttivo degli incendi che divampano nella boscaglia. Ma, paradossalmente, alcuni dei semi dispersi dalle formiche non possono germinare se non sono esposti a livelli di calore più bassi, non dannosi. Scavando nei formicai si è trovato che i semi vengono sepolti a profondità diverse. Questa disposizione può arrecare vantaggio alle piante, perché fa sì che almeno alcuni semi siano sepolti a una profondità sufficiente a proteggerli dal calore eccessivo, ma a una temperatura ancora abbastanza elevata da farli germinare.

Diversamente dagli uccelli e dai mammiferi, che disperdono ampiamente i semi nell'habitat, le formiche li portano nei loro nidi, che hanno una localizzazione ben precisa. Si tratta di un al-

Le formiche costruiscono spesso il nido su tronchi marcescenti. Le piante qui fotografate si sono sviluppate da semi portati nel nido. Quasi tutte le piante visibili nell'immagine appartengono alla specie *Carex pedunculata*, ma si notano anche viole mammole ed epatiche.

tro schema comportamentale che favorisce la sopravvivenza dei semi. Per esempio, nelle foreste moderatamente umide, le formiche costruiscono spesso i nidi nei tronchi e nei ceppi marcescenti che sporgono dal suolo. Queste posizioni sono meno suscettibili di essere sommerse dalle inondazioni primaverili, e ciò rappresenta un vantaggio sia per le formiche sia per i semi.

Le colonie di formiche, come tutte le altre società animali e anche (soprattutto) quelle umane, tendono ad accumulare materiale di scarto. I loro cumuli di rifiuti contengono resti di prede, feci, membri della colonia morti e un gran numero di altri oggetti (alcuni dei quali non sembrano avere un'utilizzazione evidente) che le formiche raccolgono e invariabilmente portano al nido. I semi germinanti e le plantule, particolarmente quelle di specie mirmecocore, possono trarre beneficio dal fatto di crescere su questi cumuli.

I rifiuti organici contengono spesso elevate concentrazioni di sostanze nutritive di importanza fondamentale per la crescita della pianta: è questo il motivo per cui chi si occupa di giardinaggio provvede sempre ad avere cumuli di concime e gli agricoltori spargono sui campi il letame che raccolgono dalle stalle. I livelli di materiale organico, e quindi di azoto, potassio e fosforo, sono spesso più elevati nei formicai che non nel terreno circostante. Pertanto i rifiuti prodotti da una colonia di formiche possono costituire una piccola, ma sempre disponibile, provvista di fertilizzante per le plantule, in grado di nutrirle nei loro primi, delicati stadi di crescita.

La sopravvivenza delle plantule può essere favorita anche da altre caratteristiche fisiche del suolo nei formicai e attorno a essi. La costruzione del formicaio rende spesso il terreno soffice e ben aerato e ne aumenta la capacità igroscopica. Alcuni ricercatori ritengono che il vantaggio fondamentale offerto da molti formicai alle plantule sia quello di garantire un rifornimento d'acqua, sia pur modesto, almeno fino a quando le radici si sviluppano a sufficienza da poter trovare l'acqua da sole.

È chiaro, quindi, che le formiche possono esercitare una profonda influenza sulle condizioni ecologiche in cui avviene lo sviluppo dei semi. Per studiare l'impatto che la mirmecocoria ha avuto sull'evoluzione è stato seguito il destino di vari semi in esperimenti sul campo. Queste ricerche si sono basate fondamentalmente su confronti tra due gruppi di semi, uno dei quali è stato trasportato dalle formiche nel nido mentre l'altro è stato piantato a mano in maniera casuale nello stesso habitat. Un esempio di esperimento di questo tipo è un vecchio studio eseguito nell'Inghilterra meridionale su due specie mirmecocore di viole. Esso ha dimostrato che dopo tre anni, quando i semi erano germinati, le sole plantule spuntate dal suolo e poi sopravvissute erano quelle che si erano sviluppate dai semi del gruppo trasportato dalle formiche nel nido.

Uno studio più recente è stato portato a termine su una pianta biennale, *Corydalis aurea*, che produce semi nel secondo anno di vita. Frances M. Hanzawa del Grinnell College ha scoperto che il tasso di sopravvivenza delle plantule cresciute nei formicai e di quelle cresciute al di fuori di essi era simile. Tuttavia un maggior numero di plantule cresciute nei formicai superava l'inverno e diventava una pianta adulta in grado di riprodursi. Pertanto vi era una differenza rilevante nel numero totale di semi prodotti dai due gruppi per una seconda generazione: le piante cresciute sui formicai fornivano un numero di semi doppio rispetto alle piante di controllo.

Poiché i due gruppi avevano esattamente lo stesso numero di semi nella prima generazione, è chiaro che una popolazione di *Corydalis aurea* cresciuta in presenza di formiche disseminatrici può svilupparsi molto più rapidamente di una popolazione che non abbia avuto questo vantaggio. E una popolazione che si accresce velocemente ha una maggior probabilità di vincere la competizione con altre piante per le sostanze nutritive, lo spazio e altre risorse. Pertanto, la ricerca di Hanzawa ha dimostrato che le condizioni ecologiche per la dispersione dei semi, per esempio la presenza di formiche, hanno influito sul potenziale evolutivo delle popolazioni vegetali.

La mirmecocoria, per quanto si è detto, presenta molti ovvi vantaggi per alcune specie di piante. Finora, però, non sappiamo esattamente come le formiche traggano vantaggio da questo tipo di interazione. Sappiamo che gli elaiosomi sono avidamente ricercati dalle operaie, staccati rapidamente con un morso dai semi e somministrati alle larve. Ma si deve ancora valutare quanto questo comportamento influisca sul tasso di crescita delle colonie di formiche.

Vale anche la pena di sottolineare che non tutte le formiche si dedicano alla dispersione dei semi. Quando i semi cadono dalla pianta, solo una frazione delle specie di formiche presenti in un habitat è attratta dagli elaiosomi. In queste formiche deve essere presente una specializzazione, ma non si sa ancora se questa sia di natura comportamentale, morfologica, alimentare o di altro tipo.

Pertanto la dispersione dei semi a opera delle formiche è un modello importante per lo studio di un'ampia gamma di interazioni pianta/animale che in un certo senso sembrano asimmetriche. È evidente che le piante si sono adattate a interagire con le formiche e il carattere adattativo più lampante è la presenza degli elaiosomi; gli adattamenti delle formiche nei riguardi delle piante sono, invece, molto meno chiari.

Come sistema di dispersione dei semi la mirmecocoria è valida ma non a tutta prova. Un elaiosoma rappresenta un pasto ambito per molti differenti tipi di formiche. Gli esperimenti con *Corydalis aurea* hanno dimostrato che nessuna plantula è mai spuntata dai nidi di certe specie di formiche. In effetti, queste formiche sembrano avere un comportamento parassitario come «rapinatrici di elaiosomi» e, nell'impossessarsi di questi ultimi, probabilmente distruggono i semi o le plantule da essi germinate.

A parte le formiche rapinatrici, una decina di altri fattori in un dato habitat può contribuire al successo o al fallimento della mirmecocoria come meccanismo di dispersione dei semi. La pioggia può allagare i formicai e le colonie di formiche possono essere soggette a epidemie di malattie fungine o a massicci assalti di predatori. Se abbondano le altre fonti alimentari, gli elaiosomi possono attrarre le formiche in misura minore; inoltre, se si instaura una competizione tra varie specie vegetali per attirare l'attenzione delle formiche, i semi dotati di elaiosomi più piccoli possono venire ignorati.

Poiché l'esito della dispersione dei semi da parte delle formiche può essere molto variabile, Hall Cushman della Macquarie University e John F. Addicott dell'Università dell'Alberta hanno proposto che la mirmecocoria sia una forma di mutualismo condizionato. In un dato momento e luogo, a seconda delle condizioni predominanti, l'interazione può avvenire in modo non ottimale. Quando invece tutte le condizioni sono in regola, i vantaggi della mirmecocoria sia per le piante sia per le formiche risultano molto significativi. Essi sono sufficientemente importanti da rafforzare in misura notevole le pressioni che tendono a conservare i caratteri che rendono possibile questo comportamento.

Via via che si allunga la lista delle piante mirmecocore note, i biologi possono sperare di comprendere sempre meglio l'importanza di questo meccanismo di dispersione dei semi in tutto il mondo. Ulteriori studi sui benefici della mirmecocoria per le piante e le formiche dovrebbero anche aiutarci a chiarire i particolari delle interazioni mutualistiche e delle loro conseguenze evolutive.

BIBLIOGRAFIA

BERG ROLF Y., *Myrmecochorous Plants in Australia and Their Dispersal by Ants* in «Australian Journal of Botany», 23, n. 3, giugno 1975.

HANDEL STEVEN N., *Dispersal Ecology of* Carex pedunculata *(Cyperaceae): A New North American Myrmecochore* in «American Journal of Botany», 63, n. 8, settembre 1976.

HANDEL STEVEN N., *The Competitive Relationship of Three Woodland Sedges and Its Bearing on the Evolution of Ant-Dispersal of* Carex pedunculata in «Evolution», 32, n. 1, marzo 1978.

BEATTIE ANDREW J., *The Evolutionary Ecology of Ant-Plant Mutualisms*, Cambridge University Press, 1985.

FENNER M., *Seed Ecology*, Routledge, Chapman & Hall, Inc., 1985.

HANZAWA FRANCES M., BEATTIE ANDREW J. e CULVER DAVID C., *Directed Dispersal: Demographic Analysis of an Ant-Seed Mutualism* in «American Naturalist», 131, n. 1, gennaio 1988.

La Terra tende a riscaldarsi?

L'analisi delle registrazioni effettuate a terra e in mare nell'arco di 300 anni conferma che il nostro pianeta si è riscaldato mediamente di mezzo grado Celsius, ma non è certo se questa tendenza perdurerà nel futuro

di Philip D. Jones e Tom M. L. Wigley

I dati meteorologici raccolti negli ultimi 100 anni fanno ritenere che la Terra si stia riscaldando. È davvero così? Molti fattori, dalle modifiche di progetto degli strumenti termometrici alla crescita delle «isole urbane di calore», possono influenzare i dati e dare l'impressione fittizia che sia in atto una tendenza al riscaldamento globale.

E ammesso che la temperatura sia realmente aumentata negli ultimi 100 anni, c'è ragione di ritenere che il riscaldamento avrà seguito? Secondo alcuni modelli al calcolatore i gas che intrappolano il calore emessi negli ultimi due secoli dovrebbero provocare nei prossimi 50-75 anni un aumento della temperatura media della Terra compreso fra uno e quattro gradi Celsius. Questi modelli, però, sono semplificazioni relativamente grossolane della miriade di complessi processi fisici che hanno luogo nell'atmosfera e negli oceani; non possono pertanto dimostrare che le emissioni di gas-serra porteranno a una alterazione significativa del clima terrestre.

Abbiamo di recente completato un'analisi decennale delle tendenze globali della temperatura, correggendo sistematicamente i dati provenienti da osservazioni condotte a terra e in mare per eliminare le possibili cause di distorsione. Il nostro lavoro dimostra in modo definitivo che il clima terrestre, per quanto soggetto ad ampie variazioni su periodi pari o inferiori a 10 anni, è diventato in generale più caldo durante l'ultimo secolo. La tendenza all'aumento della temperatura è stata interrotta da un breve intervallo durato all'incirca dal 1940 al 1970, ma da allora la curva è tornata a salire e non dà segni di cedimento.

Se la tendenza al riscaldamento globale è certa, meno certe sono le cause del fenomeno. Per quanto il riscaldamento osservato sia coerente con l'effetto serra, ci sono altri fattori, dalle eruzioni vulcaniche alle correnti oceaniche, che influenzano il clima e velano il «segnale serra». Alcune di queste ambiguità saranno forse risolte dai dati che verranno raccolti nei prossimi decenni (durante i quali è prevista una sostanziale intensificazione del riscaldamento).

Le registrazioni storiche della temperatura sono di importanza cruciale per quantificare il riscaldamento subito dai due emisferi da quando la Rivoluzione industriale ha innescato un massiccio aumento nei livelli atmosferici di anidride carbonica e degli altri gas-serra. Sfortunatamente questi dati sono di difficile reperibilità. Attualmente le misurazioni di temperatura su scala mondiale sono compilate dal World Weather Watch System, una rete di cooperazione tra servizi meteorologici nazionali. Nei secoli passati, invece, le osservazioni erano registrate soprattutto da singoli che lavoravano in modo non coordinato.

Circa 10 anni fa, spinti dalle crescenti preoccupazioni in fatto di cambiamento del clima globale, noi e i nostri colleghi della Climatic Research Unit della University of East Anglia avviammo un progetto per raccogliere e analizzare, una volta per tutte, ogni registrazione di temperatura disponibile. In questa impresa fummo sostenuti dall'US Department of Energy e dalla collaborazione di Raymond S. Bradley dell'Università del Massachusetts ad Amherst e di Henry F. Diaz dell'Environmental Resources Laboratory della National Oceanic and Atmospheric Administration (NOAA).

Il compito non era semplice. Pochi di coloro che, nei secoli XVIII e XIX, avevano allestito reti stabili di osservazione meteorologica potevano prevedere quanto i loro dati sarebbero serviti alle successive generazioni di scienziati per lo studio delle variazioni climatiche. Non c'è da stupirsi che queste registrazioni siano frammentarie, spesso incomplete e incoerenti. Ciononostante, quantificando e poi eliminando le incertezze nei dati, siamo riusciti a ricostruire in modo abbastanza preciso l'andamento del clima negli ultimi 300 anni circa, da quando hanno avuto inizio le registrazioni meteorologiche strumentali.

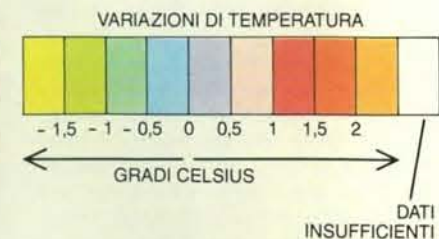
Le registrazioni più antiche sono oggi di scarsa utilità in quanto coprono solo l'Europa occidentale. Inoltre, molti dei primi dati relativi alla temperatura sono andati perduti o sopravvivono solo in forma sommaria. Ancor più si sarebbe perso se non fosse stato per gli sforzi di Heinrich Wilhelm Dove, un meteorologo tedesco che raccolse, soprattutto per corrispondenza, il maggior numero di dati possibile.

Le compilazioni di Dove sono di grande importanza, e le sue analisi dei dati erano molto avanzate per i primi decenni dell'Ottocento. Purtroppo oggi hanno un valore limitato perché non includono le zone interne di Africa, Asia, America Meridionale e Australia. A partire dal 1850, i vari enti meteorologici hanno lavorato congiuntamente per raccogliere e archiviare i dati sulla temperatura; le registrazioni successive a quella data sono quindi molto più complete.

Gradualmente si sono istituite stazioni di osservazione in quasi tutto il resto del mondo; alla fine degli anni cinquanta la rete ha raggiunto anche l'Antartide. (Vi sono state pause significative in questa diffusione: il congelamento del mercurio, per esempio, ha ostacolato le prime misurazioni nelle zone settentrionali dell'Unione Sovietica e del Canada.)

I tentativi moderni per determinare le tendenze globali della temperatura ebbero inizio circa 30 anni fa. Inizialmente, però, il lavoro era ostacolato dalla scarsità dei siti (qualche centinaio) per i quali erano stati pubblicati i dati. La nostra ricerca ha portato alla luce più di 3000 serie di registrazioni da analizzare, molte delle quali inedite.

Avendo così a disposizione un ampio elenco, si cominciò con lo sfoltirlo. Il requisito fondamentale era l'omogeneità: le registrazioni di temperatura dovevano riflettere solo variazioni meteorologiche giornaliere e variazioni climatiche di lungo termine. Le fluttuazioni attribuibili ad altre cause - come lo spostamento di una stazione di rilevamento, incoerenze nel calcolo delle temperature medie mensili oppure cambiamenti nella strumentazione, nei tempi di osservazione o nelle condizioni locali - possono inficiare seriamente l'attendibilità dei dati. I fattori ambientali sono la fonte più pericolosa di disomogeneità, in quanto possono introdurre forti distorsioni invece di limitarsi a ridurre la precisione complessiva. La crescita delle città, per esempio, crea isole urbane di calore che fanno aumentare i valori di temperatura misurati nelle stazioni vicine.

Per valutare l'omogeneità, confrontammo le registrazioni di ciascuna stazione con quelle di stazioni distanti da poche decine a qualche centinaio di chilometri. Salti o tendenze registrati da una stazione ma non dalle altre sono in genere segno di disomogeneità. Il procedimento di omogeneizzazione funziona adeguatamente salvo che in quelle regioni in cui le stazioni più vicine distano tra loro più di qualche centinaio di chilometri o in cui tutte le stazioni sono influenzate da fattori correlati, come per esem-

La tendenza al riscaldamento, come indicano i dati raccolti tra il 1967 e il 1986, varia nel mondo da regione a regione. Tanto nell'emisfero boreale (*pagina a fronte*) quanto in quello australe (*qui accanto*) la maggior parte delle regioni si è riscaldata, a volte marcatamente, ma un piccolo numero - soprattutto nelle aree settentrionali del Pacifico e dell'Atlantico - si è raffreddato. Se le previsioni sono esatte, il riscaldamento del globo si accentuerà nei prossimi decenni.

La variabilità del territorio influisce sulla lettura delle temperature producendo tendenze spurie. Nel XIX secolo (*qui sopra*) le città erano in genere relativamente piccole e avevano quindi scarso effetto sul clima delle regioni circostanti. Oggi (*pagina a fronte*), le isole urbane di calore influenzano direttamente il clima, alzando artificiosamente la temperatura misurata dalle stazioni meteorologiche vicine. L'installazione delle stazioni negli aeroporti può invece dar luogo a tendenze apparenti in senso opposto. Anche l'attendibilità dei dati presi in mare è cambiata nel tempo. Via via che le navi sono diventate più grandi, la lettura della temperatura atmosferica è stata effettuata più lontano dalla superficie dell'oceano (dove l'aria è più calda), creando una falsa tendenza al raffreddamento.

pio analoghi tassi di crescita urbana.

Sulla base di questi confronti, abbiamo convalidato i dati di alcune stazioni, ad altri abbiamo applicato adeguati fattori di correzione, e infine ne abbiamo eliminato circa il 10 per cento per l'impossibilità di correggerli in modo sicuro. Abbiamo anche escluso alcune stazioni che avevano cessato prima del 1950 di raccogliere dati. La base di dati risultante comprendeva così 1584 stazioni dell'emisfero boreale (delle 2666 registrazioni iniziali) e 293 dell'emisfero australe (delle 610 iniziali).

Partendo dai dati di queste stazioni, abbiamo dovuto calcolare le temperature medie regionali e quelle relative agli emisferi, compito più difficile di quanto possa sembrare. Una grossa fonte di errore è il ricambio delle stazioni: col passare del tempo ne sorgono di nuove e altre vengono eliminate. Se per esempio stazioni poste in luoghi relativamente più caldi (come le valli) vanno a sostituire altre situate in luoghi più freddi (come i fianchi di una collina), ovviamente sembrerà di notare una tendenza complessiva all'aumento di temperatura.

Il modo più semplice per eliminare queste tendenze spurie consiste nell'esprimere la temperatura di tutte le stazioni in termini di deviazione dalle proprie medie in un periodo di riferimento (dal 1950 al 1970) per il quale siano disponibili dati globali attendibili. (Per alcune stazioni le cui registrazioni non coprivano il periodo di riferimento, le medie furono calcolate sulla base di quelle delle stazioni vicine.)

Elaborammo poi i dati delle varie stazioni per produrre medie di area, ossia la temperatura media per ciascun punto di una griglia di cinque gradi di latitudine per 10 gradi di longitudine (circa 550 per 1100 chilometri all'equatore). Per alcune aree comprendenti diverse stazioni facemmo la media dei dati per arrivare al valore di un unico punto della griglia; in altre aree la media si basava su un'unica registrazione. I valori ottenuti servirono infine per arrivare alle temperature medie di ciascun emisfero. (Questo metodo elimina le distorsioni prodotte nella media emisferica dalle aree con molte stazioni di rilevamento.)

I nostri calcoli diedero due risultati immediati: in primo luogo, era chiaro che il clima del globo varia considerevolmente di anno in anno; in secondo luogo, avevamo la conferma di una tendenza complessiva al riscaldamento pari a mezzo grado dalla fine del XIX secolo.

Nonostante i nostri sforzi per garantire l'omogeneità dei dati analizzati, rimangono molti interrogativi sulla loro attendibilità. Le prime medie, stilate a partire da un insieme molto più piccolo di stazioni, sono comparabili con quelle di epoca più recente? Siamo sicuri di aver eliminato le distorsioni associate al riscaldamento urbano? E infine, fino a che punto temperature prese sulla terraferma rappresentano il clima di un pianeta coperto per due terzi dagli oceani?

Per stabilire se la scarsità di registrazioni più antiche rendesse inattendibili le medie relative al XIX secolo, abbiamo compilato medie corrette per un sottinsieme di stazioni, così da riprodurre la copertura più limitata del secolo scorso. L'entità delle discrepanze tra le medie del sottinsieme e i valori reali fa ritenere che le stime della temperatura anteriori al 1880 abbiano una precisione pari a solo la metà circa di quelle relative agli anni successivi al 1920. Gli stessi dati, però, mostrano che le stime della temperatura media sul periodo di un decennio - più importanti al fine di stabilire tendenze di lungo termine - sono assai precise. Ci sono dati sufficienti per individuare le medie decennali con uno scarto inferiore a 0,1 gradi Celsius a partire dal 1850 nell'emisfero boreale e dal 1880 nell'emisfero australe.

Per controllare, poi, che i nostri metodi eliminassero le distorsioni dovute alle isole urbane di calore, confrontammo le nostre medie annue per tutti gli Stati Uniti con quelle calcolate da Thomas R. Karl, del National Climatic Data Center di Asheville (North Carolina), utilizzando dati registrati soprattutto in aree rurali. I nostri dati relativi agli Stati Uniti mostrano, rispetto a quelli di Karl, un riscaldamento superiore di solo un decimo circa di grado Celsius, il che dimostra chiaramente l'efficacia della nostra accurata opera di cernita per l'eliminazione delle distorsioni urbane.

Per quanto il decimo di grado in più potesse indicare una lieve distorsione dovuta al calore urbano, poteva dipendere anche da altri fattori. Accurati confronti dei nostri dati con registrazioni in aree rurali provenienti da Unione Sovietica, Cina orientale e Australia orientale mostrano un eccesso di riscaldamento di soli 0-0,05 gradi Celsius per secolo, il che indica che fattori diversi dal calore urbano possono essere almeno parzialmente responsabili delle differenze tra le nostre medie e quelle di Karl.

Le misurazioni di temperatura di cui abbiamo parlato finora riguardano tutte le terre emerse, ossia solo un terzo circa della superficie terrestre. Ci sono però motivi per ritenere che i dati presi a terra diano buone stime delle fluttuazioni di temperatura nei due emisferi, per lo meno su periodi di decenni e di secoli. Gli strati superiori degli oceani hanno una capacità termica molto maggiore di quella dell'atmosfera e del sottile strato di crosta terrestre interessato dalle variazioni termiche nell'arco di un secolo o meno; è quindi lecito attendersi che le variazioni di temperatura relative alle terre emerse riflettano molto da vicino quelle relative agli oceani. Dato che i venti che soffiano dagli oceani verso terra e viceversa assicurano una buona comunicazione termica tra queste due parti del globo, ogni differenza tra le due dovrebbe rapidamente scomparire.

Di fatto, il forte parallelismo tra le temperature rilevate a terra e quelle prese in mare offre l'opportunità per verificare e rafforzare l'attendibilità delle stime del riscaldamento basate solo sulle registrazioni effettuate a terra. A questo proposito, la climatologia deve molto a un capitano di vascello statunitense, Matthew Fontaine Maury. Grazie alla sua attività pionieristica svolta negli anni trenta e quaranta del secolo scorso, egli diede un grande contributo alla standardizzazione dei metodi di osservazione meteorologica in mare, compresa la misurazione della temperatura dell'acqua e dell'aria. Fu in gran parte per merito del suo lavoro che nel 1853 venne firmato a Bruxelles un accordo internazionale per la raccolta e lo scambio di osservazioni meteorologiche marittime.

Da allora varie nazioni marinare - Stati Uniti, Gran Bretagna, Francia e altre - hanno conservato negli archivi giornali di bordo con osservazioni meteorologiche condotte in mare. Negli ultimi 20 anni, l'informazione contenuta in questi giornali di bordo (qualcosa come 80 milioni di osservazioni solo per quanto riguarda la temperatura superficiale del mare) è stata trasferita in due banche dati, il Comprehensive Ocean-Atmosphere Data Set compilato dalla NOAA e da altre agenzie statunitensi e una raccolta separata a cura del Meteorological Office britannico.

Come le registrazioni effettuate a terra, i dati presi in mare devono essere corretti per tener conto delle diverse tecniche di misurazione e di altri fattori. Prima del 1940, la temperatura della superficie marina era rilevata calando fuori bordo un secchio, ritirandolo pieno d'acqua in coperta e aspettando il tempo sufficiente perché il termometro segnasse la temperatura. Dai primi anni quaranta, invece, nella maggior parte dei casi la misurazione della temperatura alla superficie del mare viene effettuata nelle condotte che prelevano l'acqua marina per il raffreddamento delle macchine. Per quanto ancor oggi si facciano molte misurazioni con il secchio e anche prima del 1940 si facessero a volte misurazioni interne, il passaggio da un sistema all'altro è stato abbastanza netto e repentino.

Alcuni studi comparativi dimostrano che le misure prese in condotte interne sono di solito superiori di 0,3-0,7 gradi Celsius rispetto a quelle effettuate con un secchio in tela non isolato - un incremento approssimativamente analogo al riscaldamento riscontrato nei dati presi a terra. Questa disomogeneità va dunque corretta. I giornali di bordo, però, specificano solo dagli anni settanta se le misure sono state prese con il sistema del secchio o con quello delle condotte interne, cosicché le correzioni sono a volte problematiche.

Inoltre, non tutte le misurazioni effettuate con il secchio sono uguali. I secchi bagnati si raffreddano per evaporazione mentre vengono sollevati in coperta. L'entità esatta del raffreddamento dipende dalle condizioni meteorologiche e dalle caratteristiche isolanti del secchio. Nel tentativo di standardizzare le letture di temperatura, l'accordo di Bruxelles del 1853 specificava l'uso di secchi di legno, che hanno buone qualità isolanti, ma le flotte del XIX secolo continuarono a servirsi di secchi in tela, latta e altri materiali. Fra il 1900 e il 1940 la maggior parte delle navi era fornita di secchi in tela, che isolano poco e consentono all'acqua di raffreddarsi sensibilmente nel tragitto dalla raccolta alla misurazione. I secchi usati a partire dalla seconda guerra mondiale sono in plastica e adeguatamente isolati: le loro letture concordano con quelle fatte nelle condotte.

Anche le misurazioni della temperatura atmosferica effettuate in mare sono

soggette a problemi di omogeneità. Il più grave è dato dall'aumento, nel corso del tempo, delle dimensioni medie delle navi e di conseguenza dell'altezza del ponte (e quindi anche del termometro) rispetto al livello del mare. Di norma, la temperatura dell'aria diminuisce rapidamente con la quota, e questa maggiore altezza del ponte delle navi dovrebbe produrre una tendenza spuria al raffreddamento. Inoltre, è quasi impossibile accertare se su talune navi i termometri siano stati esposti direttamente al sole o collocati in prossimità di strutture calde.

Nelle registrazioni effettuate durante la seconda guerra mondiale si possono trovare esempi classici di diverse cause di disomogeneità. In quel periodo la maggior parte dei termometri veniva portata sulla plancia della nave, dove le letture potevano avvenire più al sicuro (ma dove la stessa struttura calda della nave alterava la temperatura atmosferica). Inoltre la grande maggioranza delle letture era effettuata di giorno, dato che era proibito accendere una luce di notte per leggere il termometro. Ne risulta che le temperature registrate durante la guerra sono sistematicamente superiori di circa un grado a quelle degli anni immediatamente precedenti o successivi.

Nelle nostre prime analisi correggevamo i dati marini confrontando le registrazioni fatte in mare in prossimità della terra con osservazioni effettuate nelle isole o nelle zone costiere. Attribuivamo le differenze tra temperature marine e terrestri alle disomogeneità delle misurazioni fatte in mare e ricavavamo fattori di correzione facendo la media di queste differenze per molte regioni. Benché questo metodo parta dal presupposto che i dati presi a terra siano omogenei, le correzioni a cui porta hanno un elevato grado di coerenza. Il necessario calcolo della media, però, limita la tecnica solo a correzioni dei dati medi annui relativi a un emisfero e non a medie locali o di scala mensile.

Abbiamo allora iniziato ad adottare una tecnica di correzione più raffinata, elaborata da Chris K. Folland and David E. Parker del Meteorological Office britannico, con la quale le misurazioni più antiche vengono modificate per tenere conto del raffreddamento per evaporazione dei secchi. L'entità del raffreddamento dipende da due fattori: le condizioni meteorologiche e il tempo di esposizione che intercorre tra il momento in cui il secchio viene portato in coperta e il momento in cui viene misurata la temperatura. Mentre le condizioni meteorologiche prevalenti dipendono dal periodo dell'anno e dal luogo in cui viene eseguita la misurazione, il tempo di esposizione è di solito ignoto e se ne deve fare una stima a partire dai dati.

Per stimare il tempo di esposizione, esprimiamo le temperature della superficie marina per un certo mese in termini della loro deviazione dalla media per quel mese nel periodo di riferimento 1950-1970. Dato che i cicli stagionali della temperatura degli oceani sono rimasti relativamente costanti negli ultimi 100 anni, queste deviazioni non dovrebbero evidenziare andamenti estate-inverno riconoscibili; la differenza tra le temperature del dicembre 1890 e la media del mese di dicembre del periodo di riferimento, per esempio, non dovrebbe essere né superiore né inferiore alla differenza tra le temperature del giugno 1890 e la media del mese di giugno del periodo di riferimento. Qualsiasi andamento stagionale dei dati può essere attribuito ai secchi, i quali si raffreddano di più o di meno a seconda del periodo dell'anno. Dalla comparsa di un andamento stagionale, quindi, possiamo stimare il tempo di esposizione e correggere i dati.

Le correzioni ricavate dai dati presi a terra e dai modelli di raffreddamento sono ampiamente concordanti nel periodo dal 1900 al 1940, quando quasi tutte le navi usavano secchi in tela. L'accordo è ancora buono per i dati del XIX secolo se si suppone che la maggior parte dei prelievi venisse eseguita con secchi in legno. Se invece si suppone che venissero usati secchi in tela, le medie relative agli emisferi dei dati presi in mare sono coerentemente di circa 0,2 gradi Celsius superiori alle medie dei dati presi a terra.

I dati di temperatura della superficie marina devono essere corretti per tener conto della tecnica di misurazione. Prima del 1940 le letture erano effettuate calando un secchio e portando l'acqua in coperta (*qui sopra*). In seguito, i termometri sono stati per lo più collocati nelle condotte di immissione dell'acqua destinata al raffreddamento delle macchine. Dal momento che l'acqua si raffredda per evaporazione, le letture fatte con il metodo del secchio possono anche essere sistematicamente inferiori di 0,7 gradi Celsius.

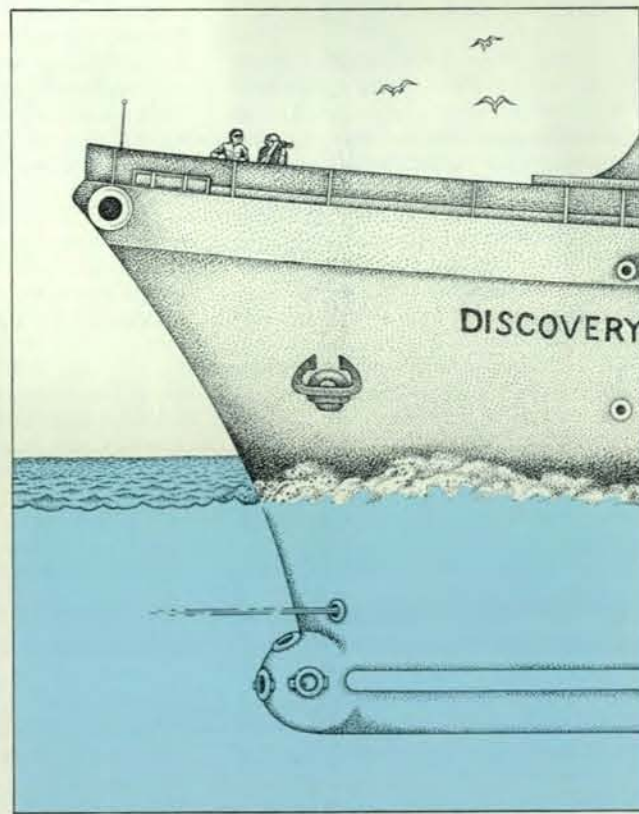
Comunque, i confronti complessivi tra dati presi a terra e in mare dimostrano un notevole parallelismo. Anche su scala temporale annua, le medie emisferiche relative alle terre emerse e agli oceani presentano una forte correlazione, e per fluttuazioni di più lungo periodo le due medie sono in accordo quasi totale. Il permanere di incertezze va addebitato alla insufficiente copertura globale. In alcune aree, in particolare nelle parti meridionali degli oceani, ancora oggi le misurazioni sono scarse. I risultati di recenti rilevamenti da satellite, però, fanno ritenere che una copertura incompleta non costituisca un serio problema. Roy W. Spencer della NOAA e John R. Christy dell'Università dell'Alabama hanno elaborato stime della temperatura per il periodo dal 1979 al 1988 partendo da dati presi dai satelliti: le loro stime concordano fortemente con le nostre serie combinate di dati presi a terra e in mare.

La variabilità climatica è evidente nelle registrazioni storiche delle medie annue e decennali della temperatura relative alle aree emerse e a quelle oceaniche dell'emisfero boreale (*in alto*), di quello australe (*al centro*) e del mondo intero (*in basso*). In tutti e tre i grafici, però, è evidente una tendenza della temperatura verso l'alto; anche gli anni più freddi dell'ultimo decennio sono stati più caldi di tutti gli anni del secolo scorso, salvo rari casi.

Dopo i 10 anni passati a rintracciare dati di temperatura e a correggerli per eliminare possibili cause di errore, possiamo affermare in modo inequivocabile che nel corso dell'ultimo secolo la temperatura del globo è aumentata. Rimangono, però, molti interrogativi: quanto è forte la tendenza al riscaldamento? Qual è la sua causa? È collegata all'effetto serra, e, se sì, perché è interrotta da un periodo di raffreddamento? E quanto è significativo il fatto che il 1987 e il 1988 siano stati i due anni più caldi tra quelli per i quali esistono dati?

A molte di queste domande si potrà dare una risposta definitiva solo con decenni di ulteriore raccolta di dati. Anche senza quei dati, comunque, gli attuali modelli di variazione del clima possono servire a una parziale spiegazione dei fenomeni. Il cambiamento del clima è determinato da fattori sia interni sia esterni; tra quelli interni si possono annoverare i cambiamenti della albedo planetaria (dovuti alle fluttuazioni naturali della copertura nuvolosa o delle caratteristiche superficiali) e degli andamenti della circolazione atmosferica od oceanica. Dato che l'andamento della circolazione atmosferica determina i flussi orizzontale e verticale di calore attraverso l'atmosfera, esso influenza il modo in cui il calore è sottratto alle terre emerse o agli oceani e a essi restituito. L'andamento della circolazione oceanica ha una forte influenza sulla temperatura degli strati bassi dell'atmosfera e sul tasso di scambio di calore tra oceani e atmosfera. Le fluttuazioni di entrambi gli andamenti possono portare a fluttuazioni di lungo termine della temperatura.

Tra i fattori esterni che influenzano il clima ve ne sono di naturali, quali la variazione di luminosità del Sole, e altri che possono essere o naturali o artificiali. Per esempio, cambiamenti nella percentuale di radiazione di piccola lunghezza d'onda che raggiunge la troposfera (lo strato di atmosfera interessato dai fenomeni meteorologici) possono derivare da un accumulo di emissioni industriali o dall'immissione naturale nella stratosfera di polvere e solfati a causa di eruzioni vulcaniche, oppure da entrambi i fattori. Le emissioni industriali possono influenzare il clima anche indirettamente, modificando l'albedo delle nubi. Concentrazioni sempre maggiori di gas-serra, a loro volta, possono incidere sul clima modificando l'assorbimento nella troposfera di radiazione di grande lunghezza d'onda in uscita.

La maggior parte dei cambiamenti climatici da un anno all'altro è dovuta a fattori interni che hanno a che fare con la circolazione dell'atmosfera. Su scale temporali più lunghe - da due a otto anni

- variazioni nella circolazione verticale degli oceani e nella temperatura della superficie marina portano a cambiamenti climatici. Per esempio, il fenomeno di El Niño (che interrompe l'andamento predominante dei venti superficiali di levante, dei venti occidentali di alta quota e della risalita di acqua fredda nel Pacifico orientale) porta a un aumento su scala mondiale delle tempeste e a una temporanea diminuzione della temperatura media del globo. In effetti, la tendenza al riscaldamento degli ultimi anni ottanta risulta ancor più significativa una volta corretta per tener conto di El Niño: il 1989 diventa l'anno più caldo mai registrato, e il 1988 e 1987 sono rispettivamente secondo e terzo.

Su scale temporali pari o superiori ai decenni ci si può attendere una variazione climatica notevole come risultato della grande inerzia termica degli oceani, che interagisce con fluttuazioni di breve periodo e accentua quelle di scala temporale più lunga. L'effetto dell'inerzia termica degli oceani può essere stimato forzando, con un disturbo casuale, un adeguato modello climatico a riprodurre le variazioni ad alta frequenza (annuale) della temperatura media globale.

Ne risulta che nel corso di un secolo le variazioni termiche a bassa frequenza possono arrivare a 0,2-0,3 gradi Celsius. In altri termini, anche il 50 per cento della tendenza osservata al riscaldamento potrebbe essere attribuito a una fluttuazione interna naturale. Un'interpretazione altrettanto plausibile, però, è che un riscaldamento molto superiore (0,7-0,8 gradi Celsius) sia parzialmente compensato da una fluttuazione di raffreddamento di origine interna.

Il clima della Terra dipende anche da fattori esterni, uno dei quali è la variabilità del Sole. Recenti osservazioni da satellite hanno confermato che l'emissione solare varia dello 0,1 per cento circa in concomitanza con il ciclo undecennale delle macchie solari - una variazione di circa 0,24 watt per metro quadrato nella radiazione che raggiunge l'alta atmosfera. Se il sistema climatico potesse rispondere immediatamente alla variazione dell'emissione solare il globo si riscalderebbe (o raffredderebbe) in misura compresa tra 0,08 e 0,24 gradi Celsius nel corso del ciclo delle macchie solari. L'inerzia termica degli oceani impedisce una risposta così rapida e l'effettiva variazione termica globale è probabilmente inferiore a 0,03 gradi Celsius.

È stata avanzata l'ipotesi che la luminosità del Sole possa variare in misura maggiore su scale temporali più lunghe. Periodi prolungati di scarsa attività delle macchie solari, come i minimi di Maunder, Spörer e Wolf (presentatisi, rispettivamente, nei periodi 1645-1715, 1450--1550 e 1280-1350) sono correlati a fasi di esteso avanzamento dei ghiacciai montani (l'ultimo fu la «piccola età glaciale», durata approssimativamente dalla metà del XVI alla metà del XVIII secolo). Si è ipotizzato che l'emissione solare sia diminuita dello 0,2-0,6 per cento durante questi eventi e che questa diminuzione sia stata responsabile dei cambiamenti climatici. Dalla fine del minimo di Maunder, comunque, non è stato osservato alcun periodo prolungato di anomala attività delle macchie solari.

Anche le modificazioni del raggio del Sole, che aumenta e diminuisce con un ciclo di circa 80 anni, possono influenzare la luminosità, ma quanto sia forte il legame tra raggio e luminosità non è chiaro. L'effetto potrebbe essere significativo o trascurabile. Le osservazioni da satellite del prossimo decennio dovrebbero risolvere questi dubbi.

Allo stadio attuale delle conoscenze, l'influenza del Sole sulle tendenze termiche del globo negli ultimi 100 anni circa è incerta, ma sembra che sia comunque piccola. Anche nelle speculazioni più ardite si prendono in considerazione fenomeni i cui effetti sono insignificanti a confronto delle emissioni di gas-serra. L'entità più probabile della diminuzione di luminosità responsabile della piccola età glaciale, anche se molto superiore a qualsiasi variazione osservata in tempi recenti, è solo di un watt per metro quadrato, equivalente a circa il 40 per cento del cambiamento avvenuto sinora per effetto serra.

Gli effetti dell'attività vulcanica sul cambiamento del clima sono meglio definiti, almeno nel breve periodo. Le eruzioni vulcaniche che immettono grandi quantità di polvere e di aerosol di solfati nella stratosfera possono provocare un significativo raffreddamento a breve termine. Pare, per esempio, che nel 1883 l'eruzione del Krakatoa, vicino a Giava, abbia raffreddato gli strati bassi dell'atmosfera di qualche decimo di grado Celsius. L'effetto ebbe inizio pochi mesi dopo l'eruzione e rimase misurabile per circa due anni. L'eruzione dell'Agung a Bali nel 1963, sebbene sia stata meno violenta e abbia immesso meno polvere nella stratosfera, ha prodotto grandi quantità di anidride solforosa e ha avuto un effetto analogo sul clima.

L'azione a lungo termine di queste eruzioni è molto più dubbia. Per quanto sembri probabile che gli aerosol vulcanici precipitino dalla stratosfera nel giro di pochi anni, si potrebbe ritenere che l'inerzia termica degli oceani moduli gli effetti delle eruzioni in modo tale da dare luogo a effetti climatici di lungo termine. Si potrebbe, per esempio, attribuire parte del riscaldamento che ha avuto luogo tra il 1920 e il 1940 all'assenza di grandi eruzioni in quel periodo.

Non esistono dati continuativi sulle concentrazioni nella stratosfera di aerosol vulcanici - in particolare solfati - ed è quindi impossibile fare una stima attendibile del loro effetto a lungo termine sul clima. Anche se ci sono numerosi dati collaterali relativi, per esempio, alle eruzioni, alla trasparenza atmosferica e alle concentrazioni di solfati nelle carote di ghiaccio della Groenlandia e dell'Antartide, le stime ricavate da queste fonti non sono reciprocamente coerenti. È difficile, quindi, trarre conclusioni inoppugnabili a proposito dell'influenza dell'attività vulcanica sul cambiamento del clima a lungo termine.

TEMPERATURE GLOBALI CORRETTE PER EL NIÑO

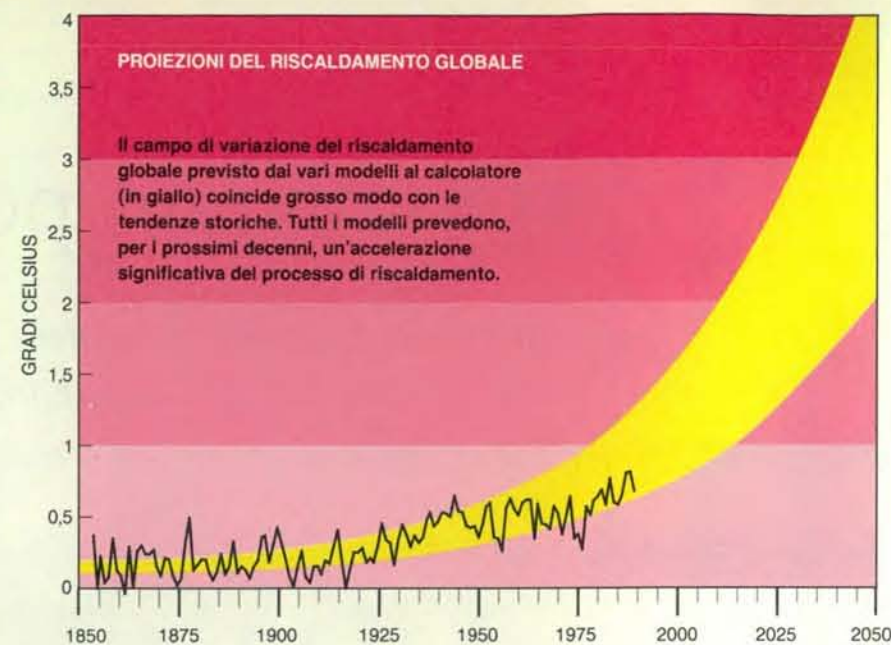
Le fluttuazioni naturali dei valori di temperatura possono mascherare il riscaldamento globale. Qui le temperature sono corrette per tener conto di El Niño, uno sconvolgimento delle consuete condizioni meteorologiche del Pacifico che fa abbassare la temperatura media della Terra. Anche altri eventi naturali, come le eruzioni vulcaniche (qui rappresentate da trattini aggiunti nella scala degli anni), possono avere effetti a breve termine di raffreddamento del clima e quindi nascondere temporaneamente il riscaldamento globale.

Che cosa si può dire allora dell'effetto serra? Almeno in questo caso esistono dati validi che documentano le variazioni della concentrazione di gas-serra nel corso degli ultimi secoli. Dal 1765 i livelli di anidride carbonica atmosferica sono aumentati da circa 280 parti per milione (ppm) a più di 350 ppm. La concentrazione di metano è più che raddoppiata, da 800 parti per miliardo (ppM) a 1700, e il protossido d'azoto è aumentato di circa il 10 per cento, da 285 ppM a 310. Negli ultimi 30 anni, la concentrazione di clorofluorocarburi è passata da una quantità sostanzialmente nulla a una parte per miliardo.

Alcuni modelli elaborati al calcolatore prevedono che il cambiamento dell'equilibrio radiativo globale provocato dall'aumento dei gas-serra è pressappoco equivalente a un aumento dell'uno per cento di luminosità del Sole. Le variazioni di livello dei gas-serra che ci sono già verificate dovrebbero finire per provocare un incremento della temperatura media del globo compreso tra 0,8 e 2,6 gradi Celsius (l'incertezza è attribuibile al fatto che, nella risposta del clima alle pressioni esterne, agiscono diversi meccanismi di retroazione non ancora chiariti).

L'inerzia termica degli oceani, comunque, impedisce al clima di reagire immediatamente alla pressione dell'effetto serra e ha ridotto il riscaldamento prevedibile durante gli ultimi 100 anni a circa 0,5-1,3 gradi Celsius. Il riscaldamento osservato, pari a circa 0,5 gradi Celsius, è quindi appena coerente con quello che ci si poteva attendere a causa dell'effetto serra.

La coerenza tra le tendenze storiche al riscaldamento e le previsioni dei modelli dell'effetto serra non significa però che quest'ultimo sia stato individuato in modo definitivo o che sia relativamente piccolo. Data l'entità della variabilità climatica naturale e degli altri fattori esterni che possono influenzare il clima, il riscaldamento osservato potrebbe essere ancora attribuibile a fattori diversi dall'effetto serra, oppure, naturalmente, un più sostanzioso riscaldamento per effetto serra avrebbe potuto essere controbilanciato da altre fluttuazioni climatiche.

Molte caratteristiche delle registrazioni storiche della temperatura, in realtà, sembrano in contrasto con l'ipotesi dell'effetto serra. Tra il 1920 e il 1940, la Terra si è riscaldata più rapidamente di quanto vorrebbero i modelli dell'effetto serra, e tra il 1940 e il 1970 si è raffreddata benché la concentrazione di gas--serra fosse in rapida crescita. Inoltre, anche i dati relativi ai due emisferi sono in contrasto con quanto previsto. Dato che l'emisfero australe ha più oceani di quello boreale, dovrebbe riscaldarsi più lentamente; invece, l'emisfero australe è risultato leggermente in testa nella corsa al riscaldamento. Queste discrepanze si possono risolvere almeno in modo qualitativo: il rapido riscaldamento della prima parte del XX secolo potrebbe essere dipeso da fattori interni o in parte dalla diminuzione dell'attività vulcanica o da variazioni dell'emissione solare, mentre il raffreddamento tra il 1940 e i primi anni settanta potrebbe essere semplicemente il risultato della variabilità naturale che ha ripreso il sopravvento sull'effetto serra.

La maggior parte dei dubbi che circondano le cause della recente variazione climatica non sarà mai risolta perché mancano i necessari dati storici. Di conseguenza è finora impossibile arrivare a una chiara e precisa interpretazione dell'innegabile riscaldamento su scala globale verificatosi in questo secolo.

Il riscaldamento osservato è all'estremo inferiore dell'intervallo previsto dai modelli dell'effetto serra, e questo può voler dire che l'effetto serra è inferiore a quanto previsto dagli attuali modelli. È anche possibile, però, che l'effetto serra sia più forte di quanto indicano i modelli e che, semplicemente, esso sia stato in parte controbilanciato da variazioni climatiche naturali o da altri fattori di origine umana.

I progressi nell'elaborazione di modelli validi e i dati che verranno raccolti nei prossimi decenni ridurranno le incertezze associate all'effetto serra e miglioreranno la previsione del futuro cambiamento del clima. Nel frattempo, però, i tentativi di spiegare le variazioni della temperatura media del globo verificatesi nel passato sono destinati al fallimento per la scarsità di informazioni sulle cause della variabilità climatica naturale.

Si tratta ovviamente di una situazione insoddisfacente per quanto riguarda i risvolti politici, ma l'incertezza non deve giustificare ritardi nella formulazione e nell'avvio di iniziative politiche volte a contenere il futuro aumento di temperatura dovuto ai gas-serra. Più tempo si aspetterà prima di agire, più forte sarà l'alterazione del clima che le generazioni future dovranno fronteggiare. Una politica di inattività sarebbe giustificabile solo se gli scienziati avessero l'assoluta certezza che l'effetto serra è trascurabile.

BIBLIOGRAFIA

WIGLEY T. M. L., ANGELL JAMES K. e JONES P. D., *Analysis of the Temperature Record* in *Detecting the Climatic Effects of Increasing Carbon Dioxide* a cura di M. C. MacCracken e F. M. Luther, US Department of Energy, Carbon Dioxide Research Division, DOE/ER-0235, dicembre 1985.

WIGLEY T. M. L., JONES P. D. e KELLY P. M., *Empirical Climate Studies: Warm World Scenarios and the Detection of Climatic Change Induced by Radiatively Active Gases* in *The Greenhouse Effect, Climatic Change, and Ecosystems* a cura di Bert Bolin, B. R. Döös, Jill Jäger e Richard A. Warrick, John Wiley & Sons, Inc., 1986.

JONES P. D., WIGLEY T. M. L. e WRIGHT P. B., *Global Temperature Variations between 1861 and 1984* in «Nature», 322, n. 6078, 31 luglio 1986.

HANSEN J. E. e LACIS A. A., *Sun and Dust versus the Greenhouse* in «Nature», in stampa.

L'arte maya in pericolo

È necessario intraprendere un'opera completa di documentazione dei dipinti murali maya, minacciati dall'inquinamento e dall'afflusso di turisti, per evitare la scomparsa definitiva di queste preziose testimonianze

di June Kinoshita

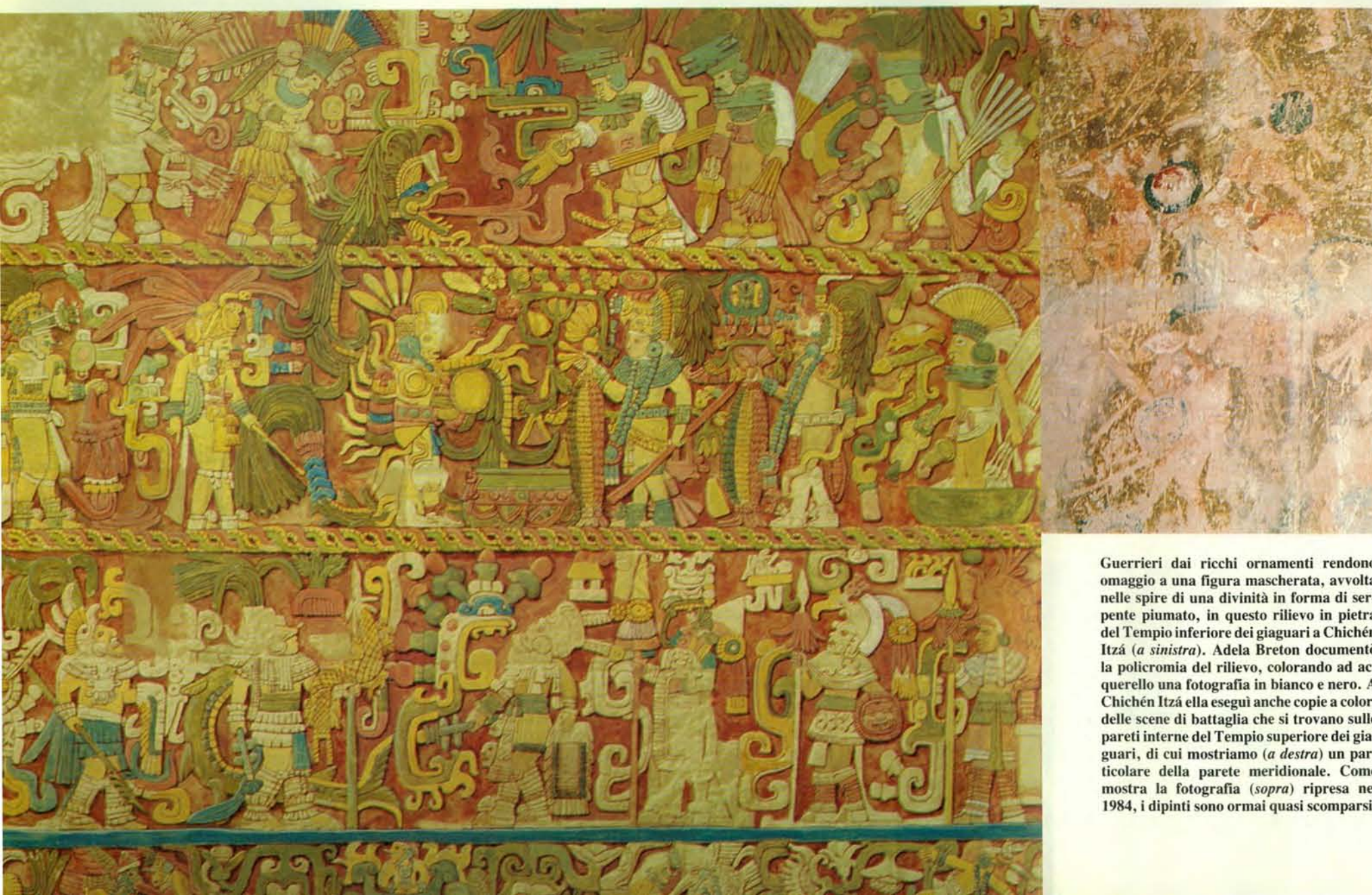
Si era nell'anno 1900 quando Adela Catherine Breton entrò per la prima volta nell'accampamento della spedizione archeologica presso il grande centro religioso maya di Chichén Itzá. Infagottata in una tenuta da cavallo vittoriana, con il volto serio, quasi da fanciulla, protetto dal sole dello Yucatán da un cappello a tesa larga, questa donna cinquantenne era senza dubbio una figura fuori dal comune. Negli otto anni successivi ella tornò più e più volte a Chichén Itzá, sfidando il caldo, i parassiti e gli intralci burocratici, per riprodurre i disegni e i colori dei rilievi in pietra e dei vivaci dipinti che ornavano le mura di calcare degli edifici in rovina.

Oggi gli studiosi della civiltà maya hanno buone ragioni per essere riconoscenti ad Adela Breton. Nei circa 80 anni trascorsi da quelle sue visite, molti dei dipinti che ella copiò con tanta cura a Chichén Itzá e in altri siti maya sono andati irrimediabilmente perduti. Il clima umido e il sole cocente della regione mesoamericana hanno danneggiato gravemente i colori, soprattutto l'azzurro brillante, a base di indaco, chiamato appunto blu maya. Vandali e turisti hanno fatto sparire molti frammenti e ciò che è rimasto è minacciato da alghe microscopiche e dalla vegetazione. In molti casi gli acquerelli di Adela Breton sono tutto ciò che resta di queste testimonianze dell'arte maya. I suoi dipinti sono stati donati al Peabody Museum della Harvard University e al City of Bristol Museum and Art Gallery, in Inghilterra, che lo scorso inverno le ha dedicato una mostra per ricordarne l'opera.

Adela Breton non fu certo in quei tempi l'unico personaggio eccentrico a mettersi in viaggio per esplorare il mondo armato di un album per disegno, ma si distinse fra gli altri per la rigorosa attenzione che usava prestare ai colori. E ciò si rivela oggi tanto più significativo in quanto, come rileva Arthur G. Miller, docente di storia dell'arte presso l'Università del Maryland a College Park, le pitture murali maya contengono una immensa quantità di informazioni espresse proprio mediante i colori.

Città splendenti di colore

Si consideri la scena di battaglia, ora gravemente danneggiata, che orna la parete meridionale del Tempio superiore dei giaguari a Chichén Itzá (900-1100 d.C.). Negli acquerelli di Adela Breton gli scudi rotondi dei combattenti appaiono ornati da un bordo rosso o blu a indicare le forze contrapposte.

Secondo George E. Stuart, archeologo della National Geographic Society, il colore era parte integrante della simbologia relativa alla visione del mondo dei popoli precolombiani. Ai punti cardina-

Guerrieri dai ricchi ornamenti rendono omaggio a una figura mascherata, avvolta nelle spire di una divinità in forma di serpente piumato, in questo rilievo in pietra del Tempio inferiore dei giaguari a Chichén Itzá (*a sinistra*). Adela Breton documentò la policromia del rilievo, colorando ad acquerello una fotografia in bianco e nero. A Chichén Itzá ella eseguì anche copie a colori delle scene di battaglia che si trovano sulle pareti interne del Tempio superiore dei giaguari, di cui mostriamo (*a destra*) un particolare della parete meridionale. Come mostra la fotografia (*sopra*) ripresa nel 1984, i dipinti sono ormai quasi scomparsi.

li, per esempio, venivano associati colori specifici: il rosso corrispondeva all'est, il nero all'ovest, il bianco al nord, il giallo al sud e il blu-verde al centro. Il colore delle vesti delle divinità indicava a quale punto cardinale esse fossero associate, mentre per quanto riguardava gli esseri umani il colore poteva esprimerne il rango oppure la provenienza. La forma degli oggetti raffigurati nelle pitture poteva essere a volte ambigua, cosicché era il colore a permettere di distinguere, per esempio, fra la giada e il mais. Anche le sculture erano in gran parte dipinte, benché oggi la pietra calcarea ci appaia nuda; ma grazie alle pitture murali possiamo avere un'idea di quale fosse all'origine l'aspetto delle decorazioni scultoree.

Non solo le sculture e i dipinti murali, ma le città intere dovevano essere uno splendore di policromie. Nel 1973 Merle Greene Robertson, direttrice del Pre-Columbian Art Research Institute di San Francisco, iniziò un'accurata indagine sui resti di pitture policrome a Palenque. Questa città del periodo classico, situata nella quasi impenetrabile foresta pluviale ai piedi di una catena di colline nel Chiapas settentrionale (Messico), raggiunse il suo apogeo durante il regno di Pacal il Grande (615-683 d.C.) ed è celebre per i delicati rilievi in stucco e le numerose iscrizioni geroglifiche.

Servendosi di tavole cromatiche di riferimento, la Robertson si diede a documentare i resti di pitture policrome che un tempo ornavano gli edifici e i bassorilievi. Secondo la sua ricostruzione, i muri erano dipinti di un rosso intenso, sia all'interno sia all'esterno degli edifici; la carnagione degli esseri umani veniva rappresentata in colore rosso, quella delle divinità in blu.

Quando la Robertson diede inizio al suo progetto gran parte delle superfici dipinte si presentava ancora pulita, ma a partire dall'inizio degli anni ottanta queste superfici cominciarono a ricoprirsi progressivamente di quella «crosta nera» che si forma quando la pietra calcarea reagisce con deposizioni umide che contengono acidi in soluzione. Le incrostazioni deturpanti finirono così con il coprire il poco colore rimasto. È da pensare che la formazione della crosta nera sia un effetto delle piogge acide, prodotte in conseguenza delle emissioni dei pozzi di petrolio e delle ciminiere industriali delle città di Coatzacoalcos e di Carmen, situate sulla costa del Golfo del Messico, 125 chilometri a nord di Palenque.

Lo scorso anno la Robertson denunciava il pericolo costituito dalle deposizioni acide in uno studio condotto per conto della National Geographic Society. Nel famoso Tempio delle iscrizioni di Palenque, la crosta nera si presentava a suo dire «così estesa e compatta che osservando [le iscrizioni] si aveva l'impressione di guardare una scultura nera e liscia». Seymour Z. Lewin, un chimico della New York University che ha esaminato le modalità di degrado in diversi siti maya, afferma che l'aspetto della crosta nera è tipico dei fenomeni di deterioramento prodotti dalle deposizioni acide, anche se i dipinti vengono danneggiati più gravemente dai microrganismi e dalla lisciviazione dei sali che non dalle deposizioni acide stesse.

Il rilevamento di Palenque a opera della Robertson era quasi al suo termine quando, nel 1982, accadde il disastro: il vulcano El Chichón, che era quiescente da 600 anni, entrò improvvisamente in attività, emettendo una densa nube di cenere e di goccioline di acido solforico. Centinaia di tonnellate di cenere corrosiva si riversarono su Palenque e ne seppellirono gli edifici come sotto una coltre di neve. Quando iniziò la stagione delle piogge la crosta nera venne asportata, ma anche il colore scomparve; sei settimane dopo l'eruzione l'intero sito sembrava ripulito, come se fosse stato tirato a lucido.

Per fortuna i rilievi eseguiti dalla Robertson erano ormai sufficientemente completi da permetterle di ricostruire l'aspetto originario degli edifici di Palenque, caratterizzati da serie di vivaci dipinti con colori vivi e in forte contrasto tra loro.

Un compito altrettanto urgente della ricostruzione della perduta policromia delle rovine è quello di costituire un *corpus* dei dipinti murali che si sono conservati fino a noi. Secondo gli esperti non sono state realizzate riproduzioni soddisfacenti neppure dei più celebri dipinti maya, quelli di Bonampak. L'esistenza di questo sito del tardo periodo classico, isolato e sperduto nella regione sudorientale dello Stato di Chiapas, venne resa nota da un esploratore inglese, Giles G. Healey, che vi fu guidato dagli indigeni nel 1946.

Il Tempio del Sole di Palenque era anticamente dipinto di rosso con ematite (un ossido di ferro) e ornato di fregi policromi, come mostra questa ricostruzione eseguita da Merle Greene Robertson (*a sinistra*). Oggi questo edificio del VII secolo d.C. si staglia bianchissimo sullo sfondo della vegetazione (*sotto*). Nella fotografia qui sopra, la Robertson confronta un campione di colore con le tracce di pigmento rimanenti sulla guancia della regina di Palenque, Zac Kuk.

Interventi controproducenti

Datate intorno all'800 d.C., le pitture di Bonampak ornano le pareti interne di un edificio in pietra a tre vani e rappresentano, con immagini realistiche e colori brillanti, una scena di battaglia, il suo esito sanguinoso e le celebrazioni dei vincitori. All'epoca della scoperta di Bonampak, era opinione diffusa tra gli studiosi che i maya, nel periodo classico, fossero una popolazione pacifica che non aveva lasciato testimonianze storiche scritte. Il sito di Bonampak fece mutare completamente questa opinione comune, rivelando che i maya erano in realtà dediti alla guerra e utilizzavano abbondantemente i geroglifici per tramandare eventi storici. La scoperta di questo sito segnò pertanto una svolta radicale nelle ricerche sulla civiltà maya.

Data l'importanza dei dipinti, le loro precarie condizioni hanno messo in allarme un gran numero di studiosi. Per un migliaio di anni il sito è stato coperto dalla fittissima foresta tropicale ed è rimasto in condizioni costanti di penombra e di elevata umidità. Subito dopo la

La fotografia riprodotta qui sopra mostra i dipinti murali di Bonampak come apparivano prima della pulitura. Le pitture erano coperte da una spessa crosta bianca di carbonato di calcio e potevano essere osservate solo inumidendo le incrostazioni con acqua o cherosene. Nobili dalle vesti bianche sono raffigurati nella fila superiore; sotto di loro, musici battono tamburi e scuotono sonagli per celebrare una vittoria in battaglia. Una magnifica replica (*sotto*) dipinta da Felipe Dávalos per il Florida Museum of Natural History di Gainesville permette di osservare nei dettagli l'abbigliamento e i geroglifici maya.

scoperta dei dipinti si abbatterono gli alberi e si realizzarono coperture di lamiera zincata; questa «sistemazione» rendeva il sito caldissimo nelle ore diurne e freddo di notte. Negli anni sessanta si tentò un intervento conservativo iniettando una soluzione di silicone nelle pareti allo scopo di consolidare la superficie dipinta, ma il silicone è stato ormai asportato e le pitture sono in condizioni peggiori di prima. Inoltre i restauratori coprirono involontariamente parte della superficie dipinta nel consolidare con malta alcune grandi fenditure nelle pareti.

Una minaccia imminente

Alcuni anni fa si è proceduto alla pulitura dei dipinti. Uno strato opaco e indurito di carbonato di calcio ne aveva ricoperto la superficie e, quando questo è stato rimosso, le scene sottostanti sono tornate alla luce in tutta la loro bellezza. Tuttavia - ironia della sorte - la pulitura potrebbe contribuire ad accelerare la distruzione dei dipinti. In effetti, essi si sono conservati fino a oggi proprio perché il carbonato di calcio fungeva in qualche modo da strato protettivo; ora che sono allo scoperto, risultano più esposti agli sbalzi di temperatura, all'umidità e ai microrganismi. Il fatto che i dipinti siano stati puliti senza provvedere a più efficaci misure di protezione è causa di serie preoccupazioni per il loro futuro.

Un guerriero dalla carnagione scura, splendidamente ornato di piume di uccello, scruta da un dipinto murale nel palazzo di Cacaxtla (*a sinistra*); una tartaruga spunta da un fregio in un'altra zona della parete (*sotto*). Si pensa che i dipinti siano stati commissionati ad artisti maya dai signori di questo sito degli olmechi-xicalanca, posto 500 chilometri a ovest del territorio maya.

Si impone pertanto la necessità di realizzare con la massima urgenza riproduzioni accurate dei dipinti di Bonampak. In realtà già parecchie ne sono state fatte, fra cui le fotografie prese da Healey nel 1946, un'altra serie di immagini fotografiche di Hans Ritter e quattro ricostruzioni pittoriche. Una di esse, una riproduzione magnificamente dettagliata dei dipinti della prima stanza, fu realizzata intorno alla metà degli anni settanta da Felipe Dávalos, per conto del Florida Museum of Natural History di Gainesville.

Tuttavia nessuna ricostruzione è completa, per quanto alcune di esse - in particolare quella di Dávalos - si possano considerare ben fatte. Quest'ultima riproduzione è anche l'unica in cui i geroglifici siano stati resi correttamente; è infatti più difficile di quanto sembri imparare a disegnare nello stile maya. Adela Breton espresse questo concetto in una lettera scritta all'insigne mayologo Alfred M. Tozzer: «Riprodurli [i dipinti maya] richiede non solo abilità artistica, ma anche la capacità molto diversa di vederli così come li vedevano gli antichi americani».

In realtà non è mai stata realizzata alcuna riproduzione fotografica o pittorica completa dei dipinti di Bonampak, ma forse si potrà rimediare presto a questa lacuna. Di recente infatti Roberto Garcia Moll, direttore dell'Istituto nazionale messicano di antropologia e storia, ha espresso interesse per un progetto in questo senso e alcuni studiosi ritengono che la documentazione di queste opere sia perfino più importante del loro restauro.

In tutta la regione mesoamericana si continuano intanto a scoprire altre pitture murali. Lo scorso anno un gruppo di archeologi messicani ha riportato alla luce una scala sepolta a Cacaxtla, un complesso palaziale in rovina situato 100 chilometri a est di Città di Messico, e ha scoperto due dipinti in condizioni di conservazione straordinarie sulle pareti che fiancheggiano i gradini. Insieme ad altri due dipinti rinvenuti a Cacaxtla intorno alla metà degli anni settanta, essi costituiscono un complesso pittorico fra i più significativi di tutta la regione mesoamericana. Secondo Ellen T. Baird dell'Università del Nebraska a Lincoln, queste pitture risalgono a un periodo compreso fra il 655 e l'835 d.C. e sono dovute agli olmechi-xicalanca, una popolazione che con tutta probabilità proveniva dalla costa del Golfo del Messico. In esse si osservano motivi tipici dell'arte del Messico centrale, ma resi in uno stile molto vicino a quello maya.

I colori dei dipinti murali maya svaniscono non appena vengono portati alla luce. Due fotografie di una scena di battaglia a Cacaxtla, una ripresa nel 1978 (*a sinistra in alto*) e l'altra nel 1989 (*a destra in alto*), mostrano il deterioramento del «blu maya» a base di indaco, mentre il rosso dell'ematite è rimasto stabile. A Tulum, parti di un dipinto sono state danneggiate dai visitatori e annerite dall'inquinamento, come documentano le due fotografie in basso, una realizzata nel 1966 (*a sinistra*) e l'altra nel 1985 (*a destra*).

Questa commistione di stili si verificò, secondo il compianto Donald Robertson della Tulane University, perché i signori di Cacaxtla affidarono l'opera ad artisti maya; questi dipinsero nello stile della loro terra, ma rappresentarono glifi caratteristici dell'altopiano messicano, in conformità al volere dei committenti. Davvero sorprendente è, nei dipinti di recente scoperta, la rappresentazione del dio del mais: pannocchie gialle recanti volti umani che spuntano tra foglie verdi su uno stelo.

Il Governo messicano ha provveduto a erigere una copertura destinata a proteggere l'intero sito. Tuttavia molti osservatori fanno notare che i dipinti scoperti alla metà degli anni settanta appaiono già sensibilmente sbiaditi. Per fortuna ne sono state fatte copie da parte di un gruppo messicano, mentre le pitture scoperte più recentemente sono state completamente fotografate lo scorso inverno sotto gli auspici della National Geographic Society. Questi dipinti sono in condizioni così buone che la documentazione fotografica può essere sufficiente, ma nel caso di opere meno ben conservate si rende necessaria la riproduzione pittorica. Anche se a molti questo metodo potrebbe apparire superato, resta il fatto che l'occhio umano è in grado di cogliere particolari che un apparecchio fotografico non riuscirebbe comunque a restituire.

Nulla di romantico

Di particolare urgenza è la documentazione di ciò che è stato scoperto, prima che scompaia definitivamente. Pochi, tuttavia, sono disposti a dedicare a questo compito il tempo e l'impegno necessari. Arthur Miller ricorda che il proprio lavoro sui dipinti di Tulum, un complesso di epoca postclassica (XV-XVI secolo) situato a dominio del mare sulla costa orientale dello Yucatán, ha richiesto quattro anni di tempo. «Occorrono dedizione, esperienza e grandissima precisione. - afferma - Non c'è nulla di romantico in tutto ciò.»

Anche Adela Breton si sarebbe detta d'accordo. Copiare i dipinti murali «mette a dura prova il cervello e i nervi, oltre che l'occhio e la mano», si lamentò con un conoscente. E di fatto è possibile che una riproduzione completa delle pitture di Bonampak non sia mai stata realizzata proprio per l'incapacità da parte degli artisti di reggere allo sforzo necessario. Le ricostruzioni realizzate da Dávalos della seconda e della terza stanza non sono neppure paragonabili a quella della prima stanza. Un'altra artista impegnata a realizzare una replica di Bonampak a Città di Messico ha iniziato dalla terza stanza, ma sembra essersi data per vinta quando si è trattato di passare alle altre due.

Il poter disporre di riproduzioni di buona qualità è sempre più importante, dal momento che molti siti maya sono seriamente minacciati dagli effetti dello sviluppo economico nella regione. A Tulum gli edifici sono anneriti dagli scarichi degli autobus; la presenza dei turisti, che pure sono necessari perché apportano valuta pregiata, provoca l'innalzamento dei livelli di umidità all'interno delle stanze in pietra e favorisce l'invasione da parte di microrganismi. I pozzi di petrolio e le ciminiere degli impianti della Pemex, la compagnia petrolifera nazionale messicana, aggravano il problema delle deposizioni acide. Di fronte a queste minacce, la soluzione migliore potrebbe essere quella di realizzare copie dei dipinti da esporre al posto degli originali, che andrebbero quindi collocati in un museo.

Anche se le opere d'arte antiche vengono protette con la massima cura possibile, il loro degrado appare inevitabile, soprattutto per quanto riguarda i dipinti. L'unico modo per conservare per il futuro queste testimonianze sembra quindi essere offerto da un completo e accurato lavoro di documentazione.

L'ANGOLO MATEMATICO

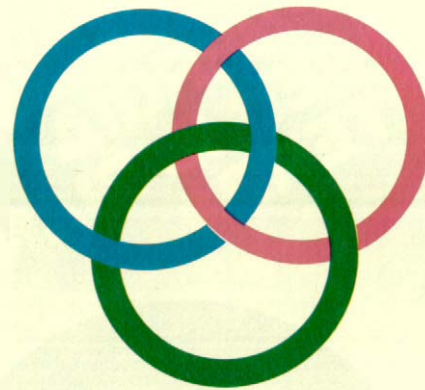
di Ian Stewart

Il polinomio di Jones scioglie l'enigma dell'universo dei nodi?

L'ingresso degli artisti era ben nascosto in un angolo, all'estremità di un androne oscuro e tutto scrostato. Quale contrasto con la brillante insegna al neon sulla facciata del teatro che mi aveva attirato in quel luogo malsano! No, non erano Mervin Snewt e il suo criceto ballerino a interessarmi (anche se la povera bestiola aveva raggiunto una fama europea grazie al satellite TV) ma un cartello collocato in basso, a destra dell'elenco dei ruoli secondari, dove delle lettere in plastica ballonzolavano appese a una guida metallica:

«Matt M. Maddox, matematico».

Ho incontrato per la prima volta Matthew Morris Maddox su un treno. È un mago specializzato in giochi di prestigio basati sulla matematica, e gli avevo promesso una visita alla prima occasione.

Lo trovai seduto al tavolo da trucco, in mezzo a una decina di tazze di cioccolata vuote, mentre giocherellava con un grande cerchio metallico. Infilò a un dito un anello che sembrava del tutto normale, lo tenne insieme al cerchio nel pugno chiuso e con l'altra mano, e fece una doppia mossa di destrezza. Poi aprì il pugno e l'anello scivolò lungo il cerchio nel quale era andato a incastrarsi.

1. *Gli anelli borromei sono allacciati ma si separano tutti quando se ne taglia uno.*

«Mi piacerebbe tanto vedere come fa» dissi.

Si girò sulla sedia:

«L'ha appena visto.»

«No, mi piacerebbe vedere qual è il trucco.»

«Trucco? Che trucco? C'è forse un trucco?»

«È evidente che non può allacciare due cappi chiusi e separati! Proprio lei, l'ultima volta che ci siamo visti, mi ha dimostrato che il numero di allacciamenti è un'invariante topologica.»

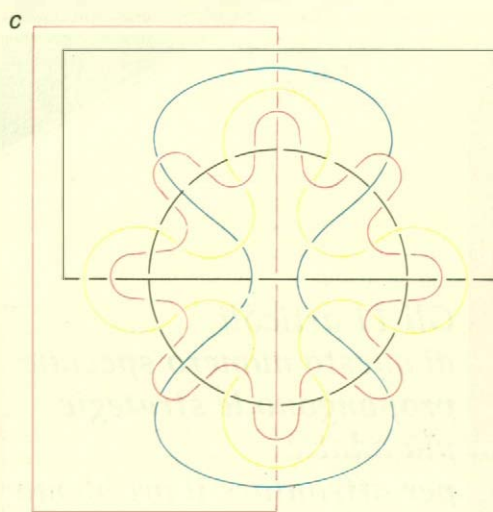
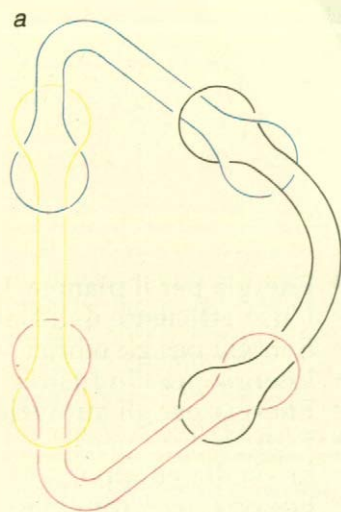
«È vero. Eppure mi ha appena visto farlo.»

È impossibile strappare qualche informazione a Maddox perché è tenuto al segreto dalle regole dell'associazione del Cerchio magico. Lasciai perdere le domande e ce ne andammo a ritemprarci nel bar all'angolo.

Mi ricordò che nel corso del nostro precedente incontro avevamo parlato degli anelli borromei (*si veda l'illustrazione 1*), un insieme di tre anelli allacciati che si separano quando se ne taglia uno qualsiasi (il nome deriva dalla famiglia Borromeo, nel cui stemma i tre anelli allacciati erano simbolo dei legami fra i tre rami della casata), e io avevo scommesso cinque dollari che non si poteva mantenere la stessa proprietà con quattro anelli.

«Non me ne sono dimenticato» dissi pagando il mio debito. «Ho fatto delle prove e ho scoperto che si può avere un allacciamento borromeo con un numero qualsiasi di anelli; anzi, esistono diverse soluzioni.

Arnaud Maes mi ha inviato tre soluzioni molto eleganti, ciascuna delle quali vale per un numero di anelli a piacere.

La prima è una serie di nodi piani allacciati (*si veda la parte* a *dell'illustrazione 2*) e la seconda è una "scala di selle" (*parte* b *dell'illustrazione 2*). Ma la terza è veramente mozzafiato. Ha letto il mio articolo sulla scoperta dei diagrammi di Venn per *n* insiemi? Arnaud Maes ha trovato il modo per convertirli in allacciamenti borromei a *n* cappi (*parte* c *dell'illustrazione 2*).»

2. *Gli «n-allacciamenti» borromei di Arnaud Maes. (a) Una catena di nodi piani. (b) Una scala di selle. (c) Soluzione basata sui diagrammi di Venn.*

Maddox era molto colpito da questa soluzione, sorprendente e di estrema bellezza, che utilizza una branca della matematica per risolvere un problema in un'altra branca del tutto diversa.

«Molto intelligente» disse. «È chiaro che ogni volta che si taglia un anello si separano tutti... Hmm! Ma come si fa a sapere che non si separano quando non viene tagliato proprio nulla?»

«Vuole che dimostri che sono davvero allacciati?»

«Proprio così.»

«Posso benissimo dimostrarlo» dissi con aria di sufficienza.

«In che modo? Non credo che possa farlo utilizzando i numeri di allacciamenti: sono tutti pari a zero perché nessuna coppia, presa singolarmente, è allacciata.»

«È vero. Bisogna usare qualcosa di meglio.»

«Che cosa, per esempio?»

Stavo quasi per dirgli che non glielo avrei rivelato; ma non voglio parlarvi per preterizione...

«I polinomi di Jones» gli confessai. Non aveva mai sentito parlare di questa strana storia.

Un allacciamento, o una catena, è un insieme composto da diversi cappi o maglie in uno spazio a tre dimensioni. I cappi possono essere ritorti o anche allacciati in modo qualsiasi (compreso quello che di solito si considera l'assenza di allacciamento). Quando c'è solo un cappio, l'allacciamento è chiamato nodo. Il problema fondamentale della teoria dei nodi è quello di distinguere un nodo dall'altro, di stabilire se due nodi sono topologicamente equivalenti, se cioè si possano ottenere uno dall'altro attraverso una serie di trasformazioni continue. In particolare vogliamo scoprire se, qualunque sia la forma di un nodo, non sia in realtà «snodato», cioè equivalente a un «non nodo». Un modo per stabilirlo è trovare delle invarianti topologiche. Esistono numeri, o oggetti matematici più complessi, che non cambiano quando il nodo è deformato in modo continuo; ne risulta che due nodi che abbiano invarianti diverse non sono mai equivalenti; tuttavia, nodi di uguale invariante possono essere equivalenti oppure non esserlo, e l'unico modo per saperlo è trovare un'equivalenza o inventare un'invariante più sensibile.

NON NODO: 1

NODO SEMPLICE A SINISTRA: $t^2 - t + 1$

OTTO: $t^2 - 3t + 1$

NODO PIANO: $(t^2 - t + 1)^2$

NODO SEMPLICE A DESTRA: $t^2 - t + 1$

NODO VACCAIO: $(t^2 - t + 1)^2$

3. *Alcuni nodi e i loro polinomi di Alexander.*

L'invariante di nodo tipica dell'era pre-Jones della teoria dei nodi era il polinomio di Alexander, elaborato nel 1926, che associa a ciascun nodo un polinomio in *t*, calcolabile secondo un procedimento standard.

Non spiegherò questo procedimento perché è simile a quello del polinomio di Jones, e quest'ultimo è migliore, ma ho indicato i polinomi di Alexander per vari nodi differenti (*si veda l'illustrazione 3*). Il polinomio di Alexander è sufficiente per distinguere un «non nodo» da un «nodo semplice», ma non va bene per distinguere un «nodo piano» da un «nodo vaccaio», un «nodo semplice a sinistra» da un «nodo semplice a destra». Prendete una cordicella e vedrete che non riuscite a trasformarli uno nell'altro e che i nodi non sono equivalenti. Il problema sta tutto nel dimostrarlo!

Tra il 1926 e il 1984, i matematici lavorarono molto su questo argomento e su problemi analoghi, e fecero anche molti progressi utilizzando metodi assai complicati. La teoria dei nodi non era affatto in un vicolo cieco, ma certo aveva bisogno di nuove prospettive.

Vaughan Jones è un matematico della Nuova Zelanda che attualmente lavora a Berkeley, in California. Nel 1984 studiava le «funzioni traccia nelle algebre degli operatori». Qualcuno notò che alcune delle sue operazioni assomigliavano a quelle che compaiono nella teoria delle trecce. Una «treccia» è un insieme aggrovigliato di corde, ma, a differenza dei nodi e delle catene, le estremità delle corde non si uniscono a formare cappi. La distinzione, comunque, non è molto importante e non impedisce di considerare una «treccia» come un tipo di allacciamento. Riflettendo sulle ragioni di una simile coincidenza, Jones scoprì che le sue funzioni traccia gli permettevano di definire un polinomio invariante per gli allacciamenti.

In un primo momento, si pensò che il polinomio di Jones fosse una variante di un polinomio di Alexander, ma ben presto ci si rese conto che era realmente qualcosa di nuovo.

Vennero trovate delle definizioni più semplici, che non facevano intervenire le algebre degli operatori, e cinque gruppi distinti di matematici scoprirono indipendentemente e contemporaneamente una generalizzazione ancora migliore

Regole relative al «polinomio di aggancio»

1 $<\bigcirc> = 1.$
2 $<E\bigcirc> = -(a^{-2} + a^{2}) <N>.$
3 $<\times> = a <\asymp> + a^{-1} <)(>.$

Più in dettaglio:

1. Il polinomio di aggancio del non nodo ○ è la costante 1.
2. Se si aggiunge un non nodo supplementare ○ a un allacciamento dato E, a cui non sia allacciato, il polinomio di aggancio è moltiplicato per $-(a^{-2} + a^{2})$.
3. Scegliamo un incrocio qualsiasi ✕ in un allacciamento e spezziamolo in due modi distinti per ottenere ≍ e)(. Così facendo otteniamo due nuovi allacciamenti che rappresentiamo temporaneamente con ≍ e)(. È possibile distinguerli: se il percorso viene da un «passaggio sotto», con una rotazione a destra sul vecchio «passaggio sotto», si ottiene ≍; se la rotazione è a sinistra si ottiene)(. Moltiplichiamo il «polinomio di aggancio» relativo al nuovo allacciamento ≍ per a, e quello relativo a)(per a^{-1}, poi sommiamo. Il risultato è il polinomio di aggancio dell'allacciamento originale.

Si noti che ≍ e)(hanno un incrocio in meno di ✕. Di conseguenza, la ripetizione di questo procedimento riduce progressivamente il numero di incroci, finché non ne rimane alcuno. Le regole 2 e 1, poi, consentono di completare il calcolo. È chiaro che il risultato è lo stesso qualunque sia l'ordine in cui vengono spezzati gli incroci.

Calcolo del «polinomio di aggancio»

Il metodo (*si veda la figura 5*) è illustrato rispetto a un nodo semplice, ma funziona allo stesso modo per qualsiasi nodo. Iniziate dal nodo semplice (1). Scegliete un incrocio (segnato da un punto giallo). Spezzatelo nei due modi possibili per ottenere (2) e (3). Notate che l'allacciamento (2) è ottenuto con una rotazione a destra partendo da un passaggio sotto verso un passaggio sopra e che il nodo (3) è ottenuto con una rotazione a sinistra. Ripetete il procedimento sugli allacciamenti (2) e (3) per ottenere gli allacciamenti (4-7) e ripetete di nuovo finché tutti gli incroci sono spezzati e rimangono solo dei non allacciamenti. Per le regole 2 e 1, il polinomio di aggancio per un non allacciamento a n componenti è $(-(a^{2} + a^{-2}))^{n-1}$. Usate questo risultato per scrivere il polinomio di aggancio per gli allacciamenti (8-15). Risalite poi utilizzando la regola 3 per calcolare il polinomio di aggancio degli allacciamenti (4-7) poi (2-3) e infine (1). Nella tabella qui sotto sono illustrati i passaggi da una tappa all'altra.

Allacciamento	Polinomio di aggancio	In che modo?
15	1	Regola 1
14	$-a^2-a^{-2}$	$-(a^2 + a^{-2}).1$ per la regola 2
13	$-a^2-a^{-2}$	$-(a^2 + a^{-2}).1$ per la regola 2
12	$a^4 + 2 + a^{-4}$	$=(-(a^2 + a^{-2}))^2$ per la regola 2, due volte
11	1	Regola 1
10	$-a^2-a^{-2}$	$-(a^2+a^{-2}).1$ per la regola 2
9	1	Regola 1
8	$-a^2-a^{-2}$	$-(a^2+a^{-2}).1$ per la regola 2
7	$-a^3$	$a<14> + a^{-1}<15>$ per la regola 3
6	$-a^3$	$a<10> + a^{-1}<11>$ per la regola 3
5	$-a^3$	$a<8> + a^{-1}<9>$ per la regola 3
4	a^5+a	$a<12> + a^{-1}<13>$ per la regola 3
3	$-a^4-a^2$	$a<6> + a^{-1}<7>$ per la regola 3
2	a^6	$a<4> + a^{-1}<5>$ per la regola 3
1	a^7-a^3-a	$a<2> + a^{-1}<3>$ per la regola 3

Quindi, il polinomio di aggancio del nodo semplice è a^7-a^3-a

per distinguere i nodi, generalizzazione alla quale viene attribuito il nome di polinomio HOMFLY (*Hoste-Ocneanu-Millett-Freyd-Lickorish-Yetter*). Oggi la decina di nuovi polinomi di nodi costituisce un vero rompicapo, perché essi non si integrano bene nel meccanismo stabilito dalla topologia algebrica. I topologi sanno calcolare questi polinomi e dimostrare teoremi al loro riguardo, ma non sanno bene che cosa siano e che cosa significhino. Sono polinomi che hanno relazioni profonde con la teoria quantistica dei campi e i topologi dovranno studiare nuovi domini matematici per capire che cosa avvenga realmente.

«I polinomi di Jones devono essere molto complicati» disse Maddox.

«Per niente! Fra un istante le farò vedere, non come calcolare un polinomio di Jones, ma come dimostrare che è un'invariante!»

Questi polinomi s
che ci si può chiedere perché non siano stati scoperti prima. Questo, comunque, non sminuisce affatto il merito di Jones: uno dei più grandi contributi che un matematico possa dare è mettere in evidenza qualcosa di semplice e contemporaneamente potente che nessuno aveva notato.

«Innanzitutto» gli dissi «studieremo il diagramma di un nodo e i movimenti di base. Il diagramma di un nodo non è altro che il disegno del nodo nel piano. A un incrocio, lei ha un passaggio sotto e un passaggio sopra. Ogni nodo è determinato dai suoi incroci e dal modo in cui i fili dei nodi sono uniti. Se deforma un nodo in modo continuo, cambia anche il suo diagramma. Lo schema degli incroci resta di solito immutato, anche se la forma dei fili che li collegano è modificata, ma non abbiamo bisogno di preoccuparci; a volte anche gli incroci stessi cambiano.

Kurt Reidemeister, uno dei primi teorici dei nodi, ha dimostrato che le uniche variazioni che si devono prendere in considerazione sono di tre tipi, che chiamerò movimenti di base.

Il primo, la torsione, prende una parte di un nodo a forma di U e la torce su se stessa per formare un cappio (*si veda la parte* a *dell'illustrazione 4*). Questo cappio può anche essere eliminato mediante un movimento contrario, una controtorsione.

Il secondo movimento di base, l'avanzamento, spinge un cappio al di sotto (o al di sopra) di un altro; il suo inverso, l'arretramento, li separa (*si veda la parte* b *dell'illustrazione 4*).

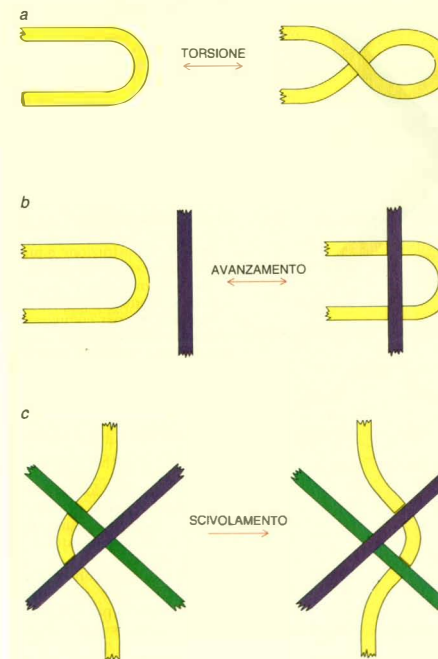
Infine, lo scivolamento sposta un filo su un incrocio formato da altri due fili (*si veda la parte* c *dell'illustrazione 4*). che l'operazione inversa è semplicemente un altro scivolamento, non avremo bisogno di parlare di "descivolamento".»

«Ogni deformazione continua di un nodo può essere rappresentata, sul suo diagramma, da una successione di movimenti di base» spiegai. «Così, per dimostrare che qualcosa è un'invariante del nodo, basta dimostrare che ciascun movimento di base lo lascia immutato. In altri termini, devo dimostrare che resta uguale dopo una torsione (o una controtorsione), un avanzamento (o un arretramento) e uno scivolamento.»

«Capisco» disse Maddox. «Allora una successione completa di movimenti di base lo lascia immutato, così come ogni deformazione continua. È la definizione stessa di invariante.»

«Precisamente! Ora le mostrerò come introdurre una piccola variante del polinomio di Jones - un metodo più semplice dovuto a Louis Kauffmann - e come utilizzare i movimenti di base per dimostrare che si tratta di un'invariante.»

Cominceremo con un oggetto un poco più semplice detto «polinomio di aggancio». Questo polinomio non è una vera invariante, ma, come vedremo tra poco, le si avvicina molto. È una combinazione di potenze positive e di potenze negative di una variabile a; per esempio:

$$a^7 - 3a^2 + 2a^{-3} - 5a^{-6}$$

è un «polinomio di aggancio». Se N è un allacciamento o un nodo, il suo polinomio di aggancio è $<N>$. Le tre regole per calcolare i polinomi di aggancio sono date nel riquadro in alto della pagina a fronte. Nel riquadro in basso, sempre nella pagina a fronte, è illustrato un calcolo tipico.

Dobbiamo controllare che il polinomio di aggancio sia invariante rispetto ai movimenti di base; in realtà non lo è. Come si vede nel riquadro a pagina 88, è invariante rispetto agli avanzamenti e agli scivolamenti, ma non rispetto alle torsioni. Una torsione moltiplica il polinomio di aggancio per a^3 o per a^{-3} secondo la direzione della torsione.

Possiamo rimediare al problema della torsione costruendo un secondo polinomio che continua a non essere modificato da avanzamenti o da scivolamenti, ma che viene diviso per a^3 o per a^{-3} da una torsione. Potremo allora moltiplicarlo ono così semplici o dividerlo per il polinomio precedente ed eliminare il problema dato dalle torsioni.

Questo secondo polinomio è più facile da calcolare. Per prima cosa, definiamo un numero detto «attorcigliamento» dell'allacciamento. Per calcolare l'attorcigliamento, è necessario per prima cosa orientare l'allacciamento, aggiungendo una freccia a ciascuna componente. Poi si definisce il valore di un incrocio : +1 o −1, come è indicato nella parte *a* dell'illustrazione 6. L'attorcigliamento $T(N)$ è la somma di tutti i valori di incrocio del nodo N (*si veda la parte* b *dell'illustrazione 6*).

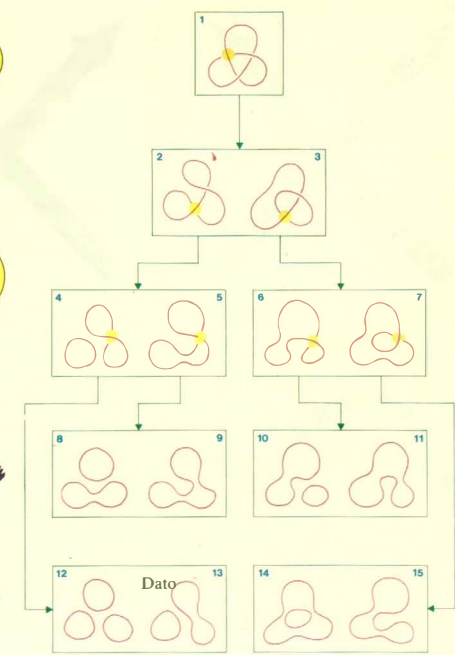
Una torsione aggiunge o toglie 1 a $T(N)$. Un «avanzamento» o uno «scivolamento» lo lasciano immutato.

Definiremo allora così il polinomio X di N:

$$X(N) = (-a)^{-3T(N)}<N>$$

La modificazione apportata agli attorcigliamenti compensa esattamente il fatto che il polinomio di aggancio sia privo di invarianza per le torsioni. Questo dimostra che $X(N)$ è un'invariante di allacciamento! Il polinomio originale di Jones è ottenuto a partire da un polinomio X sostituendo a con $t^{-1/4}$.

Per un nodo semplice orientato in un

4. *I movimenti di base: (a) la torsione, (b) l'avanzamento, (c) lo scivolamento.*

5. *Calcolo del polinomio di aggancio di un nodo semplice.*

certo modo, l'attorcigliamento è pari a 3: quindi abbiamo:

X(*nodo semplice*) =
= $(-a)^{-9}$ <nodo semplice>
= $(-a)^{-9}\,(a^7 - a^3 - a)$
= $-a^{-2} + a^{-6} + a^{-8}$

Per il «non nodo» abbiamo:

X(*non nodo*) =
= $(-a)^0$ <non nodo> = 1

Questo dimostra che il nodo semplice non è equivalente al non nodo, cioè che è veramente annodato: non può essere sciolto. Allo stesso modo, un allacciamento a due componenti (*si veda la parte a dell'illustrazione 7*) ha per «polinomio X» $-a^{10}-a^{-2}$ mentre il «non allacciamento» a due componenti (*si veda la parte* b *dell'illustrazione 7*) ha, come polinomio X, $-a^2-a^{-2}$. Quindi l'allacciamento *a* dell'illustrazione 7 non può essere «slacciato», un risultato che Maddox ha già dimostrato utilizzando i numeri di allacciamenti. [Per ricostruire la dimostrazione fatta da Maddox, seguendo i suoi stessi suggerimenti, si possono immaginare due anelli di filo variamente allacciati, supponiamo uno nero e uno rosso. Prendendo come riferimento il filo nero, si considera il numero dei passaggi «da sopra a sotto» del filo rosso e il numero dei passaggi «da sotto a sopra» e si assegna valore +1 a ciascuno dei primi e −1 a ciascuno dei secondi. Per maggiore comodità, si può immaginare l'anello nero «riempito» con una membrana: i passaggi sono allora gli attraversamenti della membrana compiuti dal filo rosso. La somma algebrica dei valori assegnati è un'invariante: se si elimina un allacciamento, si elimina una coppia di valori (+1, −1) la cui somma algebrica è nulla. Occorre notare, inoltre, che due anelli di filo non allacciati hanno, come somma algebrica dei valori dei passaggi, zero.]

Introducendo nel calcolatore il programma algebrico MAPLE, ho calcolato i polinomi X delle catene borromee di Arnaud Maes per n = 1, 2, 3, 4 (*si veda la parte* a *dell'illustrazione 2*). I risultati, riportati nel riquadro in alto della pagina a fronte, evidentemente non corrispondono ai polinomi X degli «slacciamenti» associati. Di conseguenza, il polinomio X dimostra che questi allacciamenti non possono davvero essere separati.

Cercate di calcolare i polinomi X per gli allacciamenti dell'illustrazione 8 utilizzando il procedimento descritto nel riquadro a pagina 86 in basso al fine di ottenere i loro polinomi di aggancio e usate l'attorcigliamento per compensare, come abbiamo spiegato in precedenza, la mancanza di invarianza rispetto alla torsione. Non ho calcolato i polinomi X degli altri allacciamenti borromei di Arnaud Maes né, a dire il vero, quelli degli anelli borromei originali. Se un lettore ottenesse dei risultati interessanti, mi piacerebbe leggere che cosa ha scoperto.

«Affascinante» disse Maddox «ma deve ancora mostrarmi come si distingue un nodo semplice dalla sua immagine speculare e come si riconoscono un nodo piano e un nodo vaccaio.»

«I polinomi di Jones non consentono di farlo» risposi. «Occorrono i polinomi HOMFLY, che sono ancora più potenti. I polinomi HOMFLY sono delle invarianti per i nodi orientati, ossia nodi in cui si definisce il senso di spostamento sui fili. Quando si deforma il nodo non si deve cambiare il senso della freccia. Le regole del calcolo (*si veda il riquadro in questa pagina in basso*) sono del tutto simili, ma ora ci sono due variabili, *l* e *m*.»

Utilizzando le regole del riquadro è possibile calcolare:

P (nodo semplice sin.) = $-2l^2 - l^4 + l^2m^2$.
P (nodo semplice ds.) = $-2l^{-2} - l^{-4} + l^{-2}m^2$.
P (nodo piano) =
$(-2l^2 - l^4 + l^2m^2)(-2l^{-2} - l^{-4} + l^{-2}m^2)$.
P (nodo vaccaio) =
$(-2l^2 - l^4 + l^2m^2)^2$.

Questo dimostra che i due «nodi semplici» non sono equivalenti, e che non lo sono neanche il «nodo piano» e il «nodo vaccaio». Era facile, una volta che Jones aveva indicato il metodo.

«Magnifico!» esclamò Maddox saltando d'eccitazione sulla sedia. «Il polinomio HOMFLY è un'invariante completa? Allacciamenti non equivalenti hanno in ogni caso dei polinomi HOMFLY differenti?»

«Sfortunatamente non è così» dissi. «Abbiamo mostrato nell'illustrazione 9 due allacciamenti non equivalenti che hanno polinomi HOMFLY identici. Invece il polinomio HOMFLY - e in realtà il polinomio di Jones - è forse sufficiente per distinguere ogni nodo vero da un non nodo, nel senso che nessuno conosce esempi di nodi che non siano equivalenti a non nodi e per i quali il polinomio di Jones o il polinomio HOMFLY sia uguale a 1. È stato verificato su 12 965 nodi con un massimo di 13 incroci.»

«Stupefacente!» disse Maddox.

«Noti comunque» aggiunsi «che nessuno ha dimostrato che un nodo con polinomio di Jones uguale a 1 è sciolto. È veramente un mistero.»

«Non so come utilizzare uno qualsiasi di questi ragionamenti per farne un nuovo gioco di prestigio: per quanto... » osservò Maddox profondamente concentrato. «No, non vedo proprio a che cosa possa servire. È incantevole; ma serve a qualcosa?»

«È matematica di base» dissi. «Anche se questi risultati non sono utili ora, un giorno lo saranno di sicuro. I nodi intervengono in tutti i domini: non servono soltanto per fare dei pacchetti. Comunque, la teoria ha applicazioni in fisica matematica.»

«Non è quello che intendevo. Quello che voglio io è guadagnare dei soldi» insistette Maddox.

«L'ho capito bene» dissi. «Lei sa che la molecola del DNA è costituita da due filamenti intrecciati che formano una doppia elica?»

«Naturalmente!»

«I biochimici analizzano la struttura del DNA utilizzando enzimi per tagliare a pezzi i suoi filamenti. Di solito le estremità si riuniscono e si ottiene così una molecola allacciata che si può osservare al microscopio elettronico. C'è però un problema: lo stesso allacciamento può sembrare molto diverso a seconda del modo in cui è stato deformato prima del-

Effetto del movimento di base su un polinomio di aggancio

Torsione
⟨ ⟩ = a ⟨ ⟩ + a^{-1} ⟨ ⟩
= $(-a(a^2+a^{-2})+a^{-1})$ ⟨ ⟩
= $-a^3$ ⟨ ⟩.
Allo stesso modo ⟨ ⟩ = $-a^{-3}$ ⟨ ⟩.

Avanzamento
⟨ ⟩ = a ⟨ ⟩ + a^{-1} ⟨ ⟩.
= $-a^{-2}$ ⟨ ⟩ + $a^{-1}(a$ ⟨ ⟩ + a^{-1} ⟨ ⟩)
= ⟨ ⟩.

Scivolamento
⟨ ⟩ = a ⟨ ⟩ + a^{-1} ⟨ ⟩ per la regola 3
= a ⟨ ⟩ + a^{-1} ⟨ ⟩ utilizzando due «avanzamenti»
= ⟨ ⟩ per la regola 3

6. *L'attorcigliamento:* (**a**) *incroci positivi e negativi,* (**b**) *la somma algebrica dei segni fornisce «l'attorcigliamento».*

Polinomi *X* per gli allacciamenti di Arnaud Maes

n = 1 $a^{12} + a^4 - a^2 - 1 - 2a^{-6} + a^{-8} - a^{-10} + a^{-12}$

n = 2 $-a^{22} - a^{14} + 2a^{12} + a^8 + 4a^4 - 2a^2 + 3 - 3a^{-2} + 2a^{-4} - 4a^{-6}$
$- 5a^{-10} + 2a^{-12} - 4a^{-14} + 2a^{-16} - a^{-18} + a^{-20}$

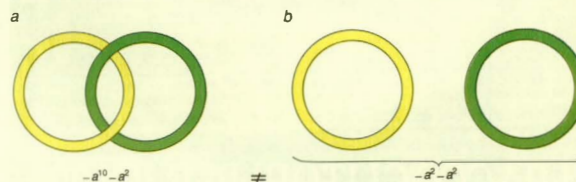
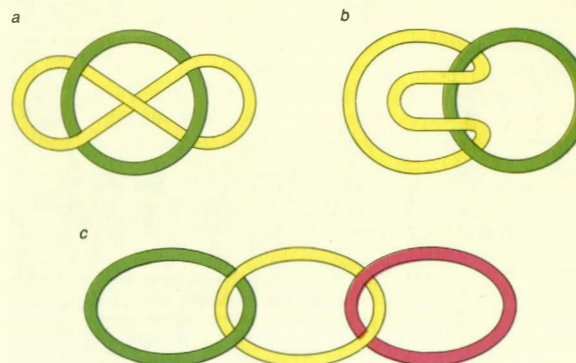
n = 3 $-a^{32} - a^{24} + 3a^{22} + 2a^{16} + 6a^{14} - 3a^{12} + 4a^{10} - 3a^8 + 3a^6$
$- 7a^4 + 10 + a^2 - 11a^{-4} + 5a^{-6} - 3a^{-8} + 11a^{-10} - a^{-12} + 13a^{-14}$
$- 3a^{-16} + 10a^{-18} - 5a^{-20} + 5a^{-22} - 3a^{-24} + a^{-26} - a^{-28}$

n = 4 $-a^{42} - a^{34} + 4a^{32} + 3a^{26} + 8a^{24} - 6a^{22} + 5a^{20} - 8a^{18} + 4a^{16}$
$- 19a^{14} + 3a^{12} - 24a^{10} + 7a^8 - 25a^6 + 20a^4 - 11a^2 + 37$
$- 8a^{-2} + 47a^{-4} - 14a^{-6} + 36a^{-8} - 28a^{-10} + 20a^{-12} - 36a^{-14}$
$+ 8a^{-16} - 36a^{-18} + 9a^{-20} - 28a^{-22} + 11a^{-24} - 16a^{-26} + 9a^{-28}$
$- 6a^{-30} + 4a^{-32} - a^{-34} + a^{-36}$

Il polinomio HOMFLY

È indicato con *P(N)*, dove *N* è un nodo orientato. Ci sono due regole:

1. P ( ) = 1
2. lP ( ) + $l^{-1}P$ ( ) + mP ( ) = 0

Nella regola 2, l'immagine mostra dei cambiamenti proprio a un incrocio del nodo. Lo stesso varrà per tutti gli incroci.

7. *Il polinomio X di un allacciamento a due componenti* (**a**) *e quello di un non allacciamento a due componenti* (**b**) *sono diversi.*

8. *Alcuni allacciamenti: sapete trovare i loro polinomi X?*

9. Un caso in cui i polinomi HOMFLY falliscono: questi due nodi differenti hanno lo stesso polinomio!

l'osservazione. Si potrebbe utilizzare il nuovo polinomio di allacciamento per identificarlo.»

«Sempre troppo poco remunerativo per Maddox!»

«Aspetti» dissi. «La American Science Foundation ha appena investito due milioni di dollari su un progetto di questo genere... Questa somma non le interesserebbe?»

Risposte

I polinomi X per gli allacciamenti dell'illustrazione 8 sono:

(a) $-2a^6 + 3a^2 - 2a^{-2} - a^{-6} - a^{10} + a^{-14}$
(b) $-a^2 - a^{-2}$ (non sono allacciati!)
(c) $a^{12} + 2a^4 + a^{-4}$.

BIBLIOGRAFIA

JONES V. F. R., *A Polynomial Invariant for Knots via von Neumann Algebras* in «Bulletin ofthe American Mathematical Society», n. 12, pp. 103-111, 1985.

LICKORISH W. B. R. e MILLETT K. C., *The New Polynomial Invariants of Knots and Links* in «Mathematics Magazine», n. 61, pp. 3-23, febbraio 1988.

STEWART IAN, *Les mathématiques*, serie *Sciences d'avenir*, Pour la Science, Parigi, 1989.